**Cross-Domain Comparison of Quantitative Technology Improvement Using Patent Derived Characteristics**

by

Christopher Lee Benson

S.B. Mechanical Engineering
Massachusetts Institute of Technology (2010)

S.M. Mechanical Engineering
S.M. Technology and Policy
Massachusetts Institute of Technology (2012)



Submitted to the Department of Mechanical Engineering
in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy in Mechanical Engineering

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2014



Signature redacted

Author..... ..................................................
Department of Mechanical Engineering
May 5th, 2014

Signature redacted

Certified by...... ..............................
Christopher L. Magee
Professor of Mechanical Engineering and Engineering Systems
Thesis Supervisor and Chair

Signature redacted

Accepted by.................................................................
David E. Hardt
Professor of Mechanical Engineering
Chairman, Committee on Graduate Students

**Cross-Domain Comparison of Quantitative Technology Improvement Using Patent Derived Characteristics**

by

Christopher Lee Benson

Submitted to the Department of Mechanical Engineering on May 5th, 2014 in partial fulfillment of the Requirements for the Degree of Doctor of Philosophy in Mechanical Engineering

# Abstract

This thesis compares the performance improvement rates of 28 technological domains with characteristics derived from the patents of the domains, seeking to objectively test theories of how and why technologies change over time. Performance metrics for 28 technological domains were tracked over time and showed exponential improvement. Each of the 28 domains increases at a different exponential technological improvement rate (the annual percentage increase in performance). These improvement rates vary substantially, including the ~36% annual improvement of Moore's Law (doubling ever 2 years) and the ~3.4% yearly improvement in electrochemical battery specific energy storage. A set of patents is selected for each domain and analyzed using patent based markers that are designed to test hypotheses of technological change. We find that the best indicator of a high improvement rate for a technology is the average number of citations that the patents in that domain receive within the first 3 years after publication, with a Pearson correlation coefficient of 0.74. This, along with several of the other tests support the hypothesis that domains whose patents are more highly cited patents are published more recently on average are likely to improve more rapidly. These measures are combined into a predictive model that can be used to accurately estimate the technological improvement rates of a domain using only patent data. A measure of reliance on basic science, the average ratio of non-patent literature citations to overall citations, did not show a correlation with improvement rate. Additionally, our data does not show a correlation between the number of patents issued and the improvement rate in a domain, however we show that patents can be used as an effort variable when compared with the functional performance metrics of a technology. By study of multiple effort variables, we find evidence to support time as the fundamental variable for which technological performance should be measured against. This is not in support of production-based theories such as Wright's Law. Ultimately the thesis provides a falsifiable quantitative and qualitative method to test how and why different technologies improve over time.

Thesis Committee:

Professor Christopher L. Magee (Chair and Supervisor)
Department of Mechanical Engineering
Engineering Systems Division

Professor A. John Hart
Department of Mechanical Engineering

Professor Steven D. Eppinger
Engineering Systems Division
Sloan School of Management

# Acknowledgements

I have to admit I never saw myself getting a PhD. The fact that I am now writing this is a testament to the wonderful intellectually curious environment, students and faculty at MIT, the support of my fantastic friends and family, and the unparalleled working relationship and friendship that Chris Magee and I have forged while completing this research. Of course most of the credit must go to my (soon-to-be) bride – Katy, for her never-ending love and emotional support have provided with the fuel for the journey.

*There are always aspects of economic life that are left out of any simplified model. There will, therefore, be problems on which it throws no light at all; worse yet, there may be problems on which it appears to throw light, but on which it actually propagates error.*

*-Robert Solow, 1974, 'Growth Theory'*

# Contents

# Chapter 1: Introduction

As technologies continue to improve at an exponential rate, there becomes an ever-greater need for understanding how technology has and will evolve (Koh and Magee, 2006). While it may be nearly impossible to fully predict how technology will change, even modest improvements in our ability to understand and potentially forecast technological change could create considerable impact in a number of areas where reducing the uncertainty of future technological capabilities is advantageous. In this chapter, we will discuss three areas that can benefit from improved understanding of technological improvement. Then we will describe in detail the problem that we are addressing in this research followed by the structure of the thesis.

## 1.1. Reducing Technological Uncertainty

Understanding how technology changes over time and what capabilities are likely to exist in several years can influence how products are designed. As an example, Schaller (1997) points out that once software designers became aware of Moore's law and the rapid exponential improvement rate of computer processors, they began to push the limits of software programs at a similar pace, increasing the lines of code in software such as Microsoft Word from 27000 in its initial release (in 1983) to 2.5 million lines in 1997.

Another area where a better understanding of technological growth could make a large impact is in private investing. Fehrenbacher (2012) writes that 'One of the key misplaced assumptions that [Silicon] Valley venture capitalists made in cleantech boom times is that the rapid progress of Moore's Law could be created for cleantech with a little bit of VC funding and Valley smarts.' In the end the technological domains associated with clean energy never matched the high technological improvement rate of computer processors, resulting in high profile failures of heavily funded companies (Worstall, 2012).

The rapid advancement of technology also affects decisions to be made in public policy. Rycroft (2006) notes that there is a 'growing divergence between time cycles of government and those of technological change... either [the government] lives with a shorter response time and run the risk of ill-considered actions or else government become less relevant'. With a greater understanding of how technologies change over time, the government will be better suited to identify technologies with high potential and allocate resources appropriately to support the development of future technologies. This knowledge could be particularly useful for funding agencies such as the National Science Foundation (NSF), National Institute of Health (NIH), Department of Energy (DOE), Department of Defense (DOD) and many other organizations that are responsible for ensuring that the United States is technologically prepared for the future. Agencies with similar goals exist around the world, such as the Campus for Research Excellence and Creative Enterprise (CREATE) in Singapore, and the European Research Commission (ERC) in Europe.

## 1.2. Problem Statement

Much of the prior work to understand how technology changes over time has been focused around case studies. Quantitative data is sometimes an important part of the case study but usually the understanding or explanation is based upon narrative. The resulting qualitative theories include the linear model of innovation, the theory of radical inventions, the theory of disruptive innovations, life-cycle theory, S-curve theory, punctuated equilibrium and combinatorial knowledge-based innovation. This thesis complements the prior work by creating quantifiable hypotheses to test such theories of technological improvement. Specifically, 28 different technical domains such as solar photovoltaics (PV), computed tomography, combustion engines, etc., are studied to determine their technological improvement rates (TIRs). Using the TIRs of each domain as the dependent variable of interest, data from patents are used to derive characteristics of each domain that are used to test hypotheses derived from the prior theories of technological change.

This thesis does not aspire to provide an answer to all of the questions about how or why technologies change. Rather, it is an attempt to define the technological improvement rates of a fairly large set of important technological fields and to better understand why these fields improve at different rates. At a higher level, this research aims to provide an avenue for future research into technological change that is based upon falsifiable tests.

## 1.3. Falsifiability

First, a quick note on the general tone of this thesis; an overarching theme of this thesis will be the falsifiability of the prior theories of technological change. The methods and results contained hereafter include contributions to their respective fields, and they all follow the common goal of creating an objective and repeatable methodology for testing theories of how technologies change over time. This theme of falsifiability is drawn largely from Popper (1962):

*The way in which knowledge progresses, and especially our scientific knowledge, is by unjustified (and unjustifiable) anticipations, by guesses, by tentative solutions to our problems, by conjectures. These conjectures are controlled by criticism; that is, by attempted refutations, which include severely critical tests. They may survive these tests; but they can never be positively justified: they can be established neither as certainly true nor even as 'probable' (in the sense of the probability calculus). Criticisms of our conjectures is of decisive importance: by bring out our mistakes it makes us understand the difficulties of the problem which we are trying to solve. This is how we become acquainted with the problem, and able to propose more mature solutions: the very refutation of a theory - that is, of any serious tentative solution to our problem - is always a step forward that takes us nearer to the truth. And this how we can learn from our mistakes.*

*Those among our theories which turn out to be highly resistant to criticism, and which appear to us at a certain moment of time to be better approximations to the truth than the other known theories may be described,* *together with the reports of their tests, as 'the science' of that time. Since none them can be positively justified, it is essentially their critical and progressive character - the fact that we can argue about their claim to solve our problems better than their competitors - which constitutes the rationality of science. (Popper, 1962)*

Many parts of this thesis are designed to be falsifiable – it is *possible* to test and then logically assert that they are false. This characteristic is important in that it allows for future testing and building upon the work done here yet requires more specificity in the details of the methods and results of the research.

## 1.4. Thesis Structure

In Chapter 2 previous theories regarding technological development are discussed. Section 2.1 explores the ideas from Wright, Moore, Ayres and Foster and related theories on *how* technologies change over time and what effort variables they should be compared against (i.e. Time, R&D, revenue). Section 2.2 introduces the works of Bush, Arrow, Dosi, Shumpeter, Christensen, Abernathy and Utterback regarding theories on *why* technologies change over time. Section 2.3 introduces patents as a proxy for inventions and prior studies from Trajtenberg, Jaffe and Henderson using patents to test technological and economic theories are discussed.

In Chapter 3 the major components of the methodology for comparing technological improvement rates with patent derived characteristics are documented. Section 3.1 discusses the definition of a technological domain (TD) and introduces the 28 domains that are analyzed in this thesis. Section 3.2 walks through the process of determining an appropriate functional performance metric (FPM) and finding the technological improvement rate (TIR) for each of the 28 TDs. Section 3.3 is a case study of 4 manufacturing technological domains on how to define appropriate FPMs and calculate TIRs. Section 3.4 introduces the classification Overlap Method

(COM) for objectively and repeatably locating a set of patents that represents a TD. Section 3.5 derives the theoretical basis for each of the main 5 hypotheses that are tested in this thesis and also the domain patent markers (DPMs) that are used to test each hypothesis. Section 3.6 provides a summary of one of the author's recent publications focused on two pairs of renewable energy technologies, and demonstrates the process of comparing the DPMs and the TIRs. Section 3.7 discusses the statistical comparison of the TIRs and the DPMs for each TD.

In Chapter 4 the results of the cross-domain TIR and DPM comparison are revealed. Section 4.1 discussed the TIRs that were derived for each of the 28 TDs and the statistical filters used to select the most complete and reliable TIR for each domain. Section 4.2 demonstrates the broad applicability of the COM and shows the patent sets that were selected to represent each of the 28 TDs. Section 4.3 walks through each of the 5 hypotheses and shows the results of the correlation tests between the TIRs and FPMs for each of the 28 TDs. Section 4.4 demonstrates using patents, revenue and R&D spending as effort variables for comparing with increasing technological capability. Section 4.5 concludes the chapter with a short summary of the most relevant results.

Chapter 5 discusses the interpretations of results of the thesis and their contributions. Section 5.1 discusses how the objective and repeatable methodology created in the thesis can contribute to further testing and understanding of theories of technological change. Section 5.2 discusses how the results of the comparison between the TIRs and DPMs contribute to theories on technological change. Section 5.3 discusses the practical implications of the results, especially how the prediction of TIRs using only patent data can be immediately impactful. Section 6 concludes the thesis with a short summary, limitations of the study, and possibilities for future work on this topic.

Due to the large breadth of the cross-domain technological improvement rate comparison with patent derived characteristics, many of the specific details about each of the 28 domains can be found in the appendices. Appendix A shows the technological improvement rates along with statistical information for each of the domains. Appendix B shows how each of the patent sets for the 28 domains was found. Appendix C shows every one of the domain patent markers tested along with their correlation to the technological improvement rates.

### 1.4.1. Acronyms and Terms

The outline above contains a number of acronyms that will be used throughout the thesis. They are summarized here for easy reference.

TD = Technological Domain: The set of artifacts that fulfill a specific generic function utilizing a particular, recognizable body of knowledge.

FPM = Functional Performance Metric: The metric that is used to evaluate the performance of a specific technological domain, which includes both measures of value and measures of cost to the consumer of a technology.

DMP = Domain Metric Pair: Specific combination of a technological domain and functional performance metric. Because each domain can be measured by several FPMs, their may be several domain-metric pairs for each domain, the number of DMPs for each domain is necessarily equal to the number of FPMs that can be used to measure performance in a domain.

TIC = Technological Improvement *Curve*: The trend/curve of improving performance for a particular DMP.

TIR = Technological Improvement *Rate*: The exponential regression coefficient derived from the TIC. This will act as the dependent variable for the large cross-domain experiment.

PRM= Point Removal Method: A method used to test the robustness of a TIR to missing points.

COM = Classification Overlap Method: The method that is used in this thesis to select relevant and complete sets of patents to represent a technological domain.

MPR = Mean Precision and Recall: A measure of how closely related a particular patent class (US or International) is to a set of search terms. This is used in the COM.

DPM = Domain Patent Markers: These are the algorithms that are applied to the patent sets for each domain in order to better understand the characteristics of the domain. These are the independent variables in the large cross-domain experiment.

# Chapter 2: Prior Literature

In this chapter the theories of how and why technologies change over time will be explored. Most of the work in this area has been published since the start of the 20th century, with many of the most significant theories originating many decades ago. The result is an ever-evolving narrative of how technologies change over time, with several theories being refined and re-worded over the course of the last century.

This section will first review theories on *how* technologies change over time and the ways in which this change can be measured. Next, theories regarding *why* technologies change and proposed explanations to the increases in technological performance over time will be explored. The combination of these two kinds of explanations forms the basis of the hypotheses that are tested in the thesis.

The final portion of the literature review will delve into more recent research on *tools* for studying technological change. Many of these tools are based upon patents and are enabled by the structured organization of the patent system and the data-analysis capabilities enabled by computers.

## 2.1. How Technologies Change over Time

In order to understand what factors may cause technologies to change over time, it is important to have an understanding of *how* technologies change over time. This section will provide a review of some of the most established methods of characterizing improvements in technology. Within each of these methods, many domains have been considered and the technological capability (e.g. cost of energy produced) has been graphed against an independent variable such as time, cumulative production, production rate or research and development (R&D) spending. Ultimately, while the effort variables used measuring the improvement of technology have varied greatly, the results been shown to be nearly equal (Nagy et al, 2013). Additionally, regardless of the effort variable, each of the methods shows a variation in rates of improvement between technological domains, which is the primary focus of this thesis.

### 2.1.1. Moore's Law and time based technological improvement

One of the most famous examples of measuring technological progress came from Moore (1965) in his seminal paper that described improvement in the manufacturing of integrated circuits. Moore recognized a trend in the ability to manufacture a higher number of components on a single manufacturing die, something that he observed to double every 18 months (revised in 1975 to every 2 years). This temporal relationship has roughly held true for the last 5 decades (Moore, 2006) as shown in figure 1.

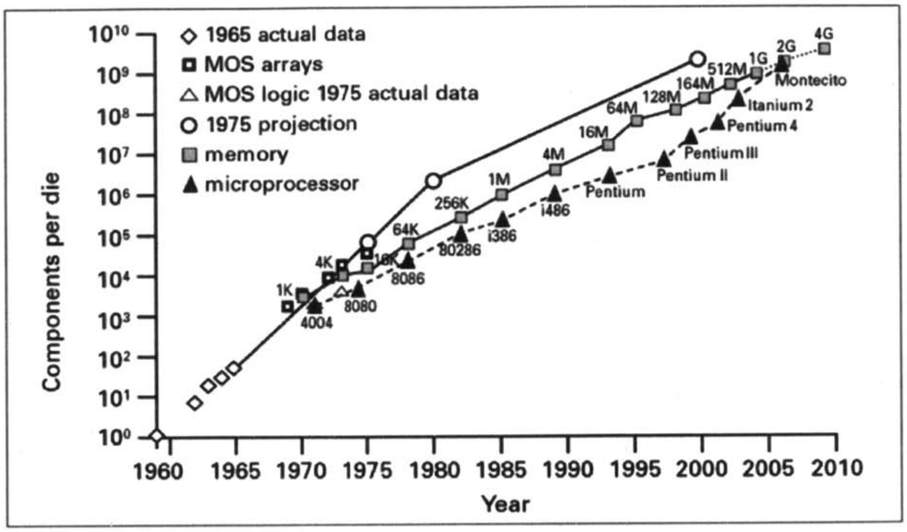



**Figure 1: The improvements in the number of components per semiconductor die for integrated circuits, adapted from Moore (2006)**

In his 2006 paper, Moore discusses the rationale and the limitations of his temporally related exponential improvement function. Figure 2 shows a summary of the rationale of the factors that have contributed to the improvement in number of transistors per die.

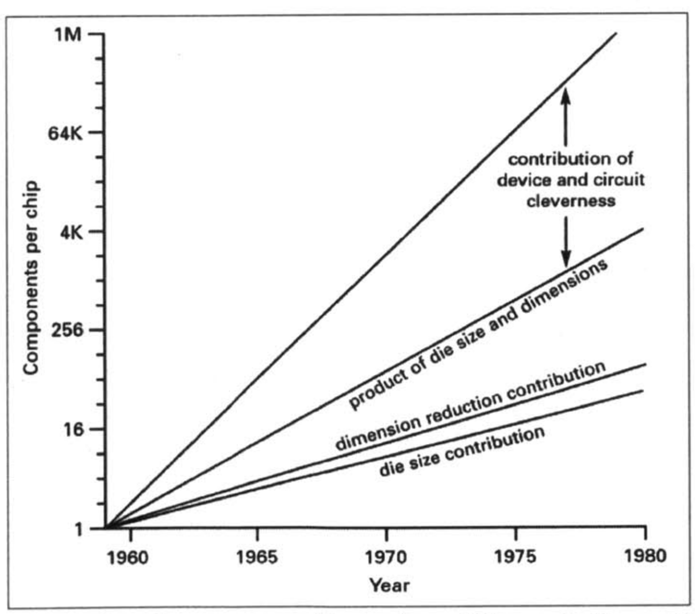



**Figure 2: Factors that have contributed to Moore's Law (Moore, 2006)**

While Moore recognizes that his law has been widely accepted and influential in the industry, he also takes time to point out that he believes it is dangerous to extrapolate exponentials. In a speech he gave in 1975, Moore pointed out that the size of the wafers was growing exponentially as well, and that this size could be expected to continue to grow. The error of this predication was later pointed out by a colleague, who stated that the wafer size would have been 57 inches in the year 2000; the colleague jokingly produced a digitally manipulated picture (Figure 3) to illustrate this. Moore uses this example as a cautionary tale to those who intend on extrapolating past performance into future expectations.

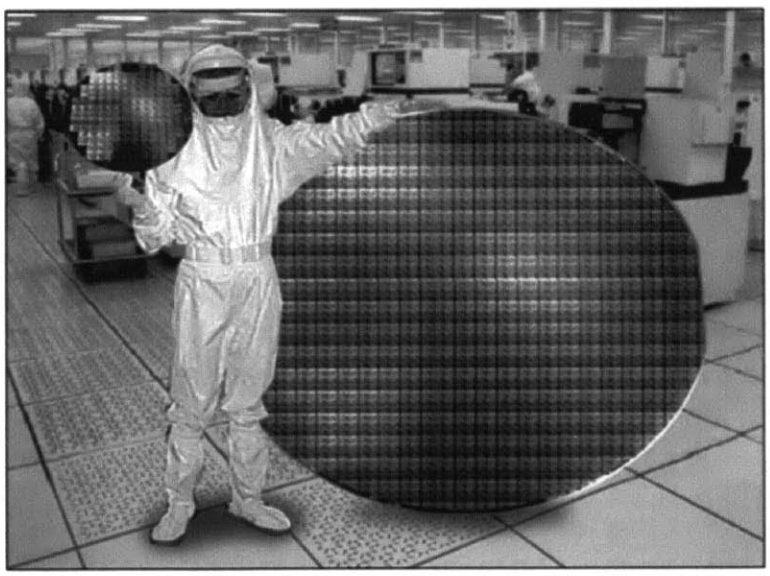


**Figure 3: Digitally manipulated photograph showing a 57 inch wafer (Moore, 2006)**

While Moore popularized the time-based exponential relationship with technological improvement, there have been many others who have found similar relationships in other industries (Martino, 1971;Nordhaus, 2009). Recently, work done by Koh and Magee (2006,

2008) has shown similar time based exponential improvements for fields such as information transmission, information storage and energy storage. The technological improvement rates within these fields have varied drastically from doubling every 2 years (~35% improvement rate) to doubling every 17 years (~4% improvement rate). These improvement rates can be modeled by equation 1, which relates the performance of a technology P to the technological improvement rate, k, and time, t. The variable B represents a scaling factor with the same units as P and relatively unimportant to our efforts which focus on k with dimensions of inverse time. The units of P are specific to each measure of technological performance and will be discussed further in section 3.2.

$$P = Be^{kt} \quad \text{(Equation 1)}$$

The time-based relationship has been confirmed by a number of authors (Magee and Devezas, 2011; Seebregts et al, 2000) to include tens of technical domains including information technology (Koh and Magee, 2006), residential lighting (Nordhaus, 1996), solar photovoltaics (Benson and Magee, 2012) and many others (Koh and Magee, 2008; Nagy et al, 2013).

### 2.1.2. Wright's Law on cumulative production related technological improvement

Another way to describe trends in the improvements of technology is to relate the amount of production of a technology to its performance. This method is accredited to being introduced by Wright (1936) in his paper 'Factors Affecting the Cost of Airplanes'. In his work, Wright noticed that when more planes were produced, the average cost per plane went down considerably. He compared the logarithm of the cost per plane with the logarithm of the number of planes produced. The resulting relationship has since been named "Wright's Law" as was graphically represented by Wright and is shown in Figure 4.

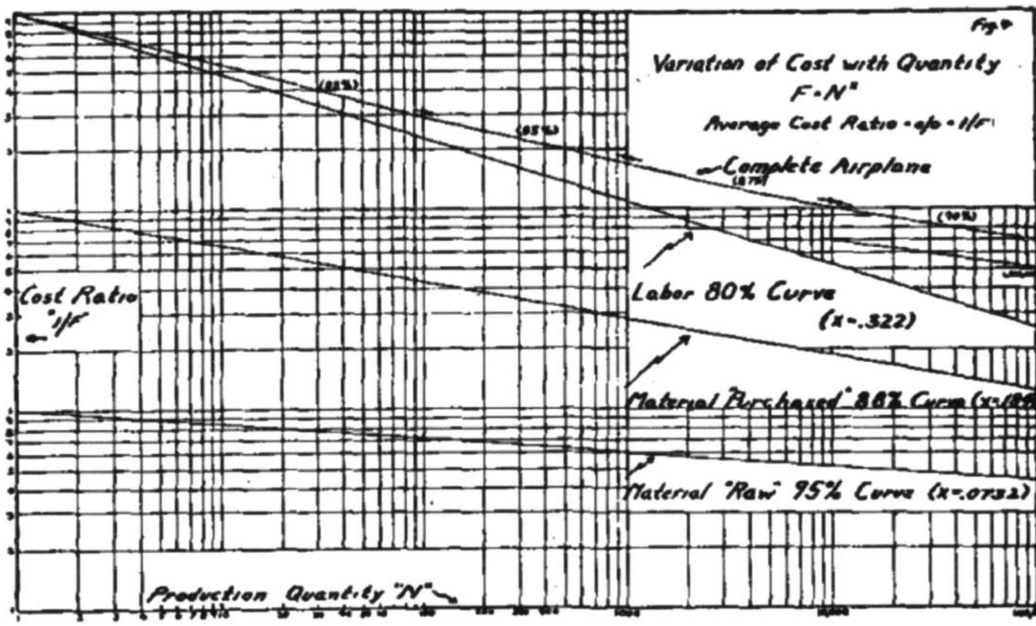



**Figure 4: Variation of Airplane Cost with Quantity Produced, adapted from Wright (1936). This figure provides a graphical representation of the reduction in cost associated with increasing production on a log –log plot.**

One of the major insights that Wright discovered was that 'the chief gain in reducing production costs was in the use of better tools and fixtures rather than anything inherent in the construction.' Wright goes on to explore the hypothetical question if it would be possible to produce a plane for $700. In order to answer this, he takes the then current price of an airplane of $18,000 when 25 are produced, and scales this to find that even when 1,000,000 planes are produced, the price only drops to $2,090 per plane. This relationship is captured in equation 2 where $Y$ is the unit cost of production, B is a scaling factor (often the cost of producing the first unit), N is the cumulative production, and x is the learning index. This effort-based relationship is explored further in section 4.4, where x is referred to as alpha.

$$Y = BN^{x} \quad \text{(Equation 2)}$$

This learning index is often converted into a learning rate *r*, which describes the reduction in unit cost of a technology with the doubling of the production volume. The learning rate can be calculated from the learning index by equation 3, which is adapted from Yelle (1979). A basic way of understanding this relationship is that a doubling in production volume will result in a unit cost that is r% times the current unit cost.

$$x = \frac{\log(r)}{\log(2)} \quad \text{(Equation 3)}$$

This learning rate, was expanded upon by Yelle (1979) who provides a comprehensive review of the work done up to that date to understand the different aspects of how technological improvements relate to cumulative production. He describes the different mathematical relations that have been suggested to relate cumulative production and unit cost, including the log-linear model (Wright's Law), the plateau model, the Stanford-B model, the DeJong Model, and the S-Model, all of which are shown in figure 5, and described in more detail by Carlson (1973; 1976). It is important to note that Wright's original formulation is in fact a power law and nearly all of the figures in Yelle's paper are Log-Log as well, thus it is probably more appropriate to refer to Wright's Law as the Log-Log model. Ultimately Yelle concedes that while there are many adaptations of the original formulation of Wright's Law that may be slightly more accurate in specific situations, the wide scope of use of the original log-log formulation and the common theoretical underpinnings of the various models allow for the sole focus of his study to remain on Wright's original law.

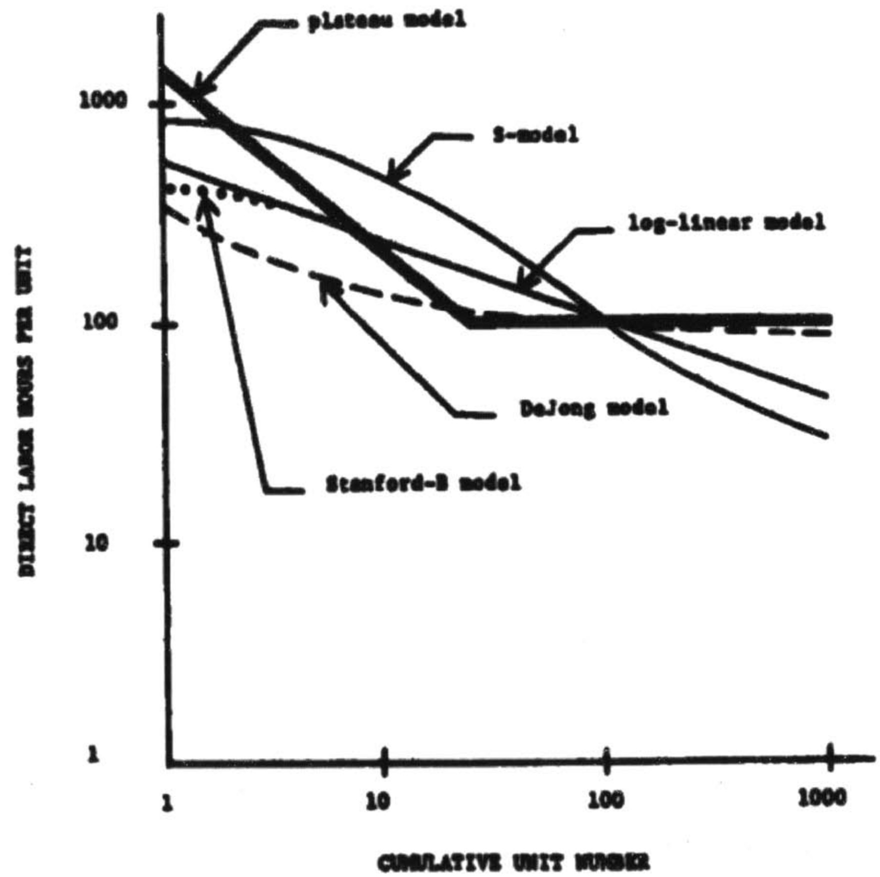



**Figure 5: The different types of curves suggested for the relation between cumulative production and unit cost (Yelle, 1979)**

While much of the work that is described by Yelle and Wright is focused on a single firm or product, the concept has since been expanded greatly to broaden the applicability to more domains and industrial scopes (Dutton and Thomas, 1984; Muth; 1986; Auerswald et al, 2000; Nemet, 2006; McNerney et al, 2011).

### 2.1.3. Other ways of measuring technological change

While Moore's Law and Wright's Law are two of the most prevalent methods of describing technological progress, there have been a number of other attempts at describing how technology improves (Nagy et al, 2013), one of which will be discussed in this section.

Many of these other methods are combinations or modifications of the first two. Goddard (1982) describes his method for understanding technological progress in the 'Opportunity Curve' and he relates technical capability to the annual *rate* of production for a particular technology. The reasoning behind the opportunity curve is that the 'growth in demand creates the opportunities and provides the funds for increased mechanization and larger batch facilities in process dominated manufacture' and that most progress is essentially driven by increasing economies of scale. Equation 4 describes Goddard's relationship using Y as the unit cost, B as a scaling factor, $N_t$ as the yearly production *rate*, and s is the associated rate of improvement between annual production rate and unit cost.

$$Y = BN_t^s \qquad \text{(Equation 4)}$$

Goddard goes on to explain how his opportunity curve relates to the other methods that have been mentioned above. Figure 6 shows how the U.S. production of silicon integrated circuits can be plotted in the several different manners described above.

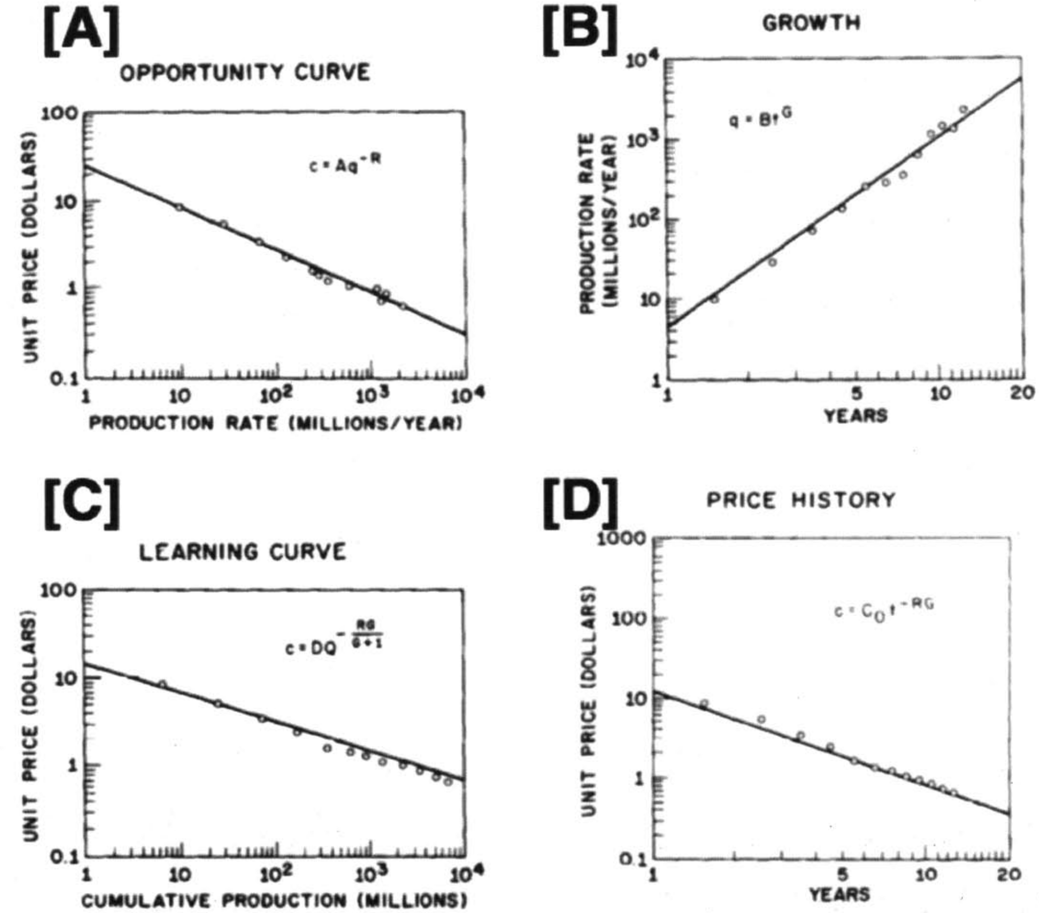



**Figure 6: Comparison of methods used to measure the production and cost of U.S. Integrated Circuits, adapted from Goddard (1982)**

Graph [A] in figure 6 shows Goddard's relationship between yearly production rate and the unit price of integrated circuits, which he uses as a measure of performance. Graph [B] shows the production rate over time, [C] shows the same data within Wright's framework and [D] shows the logarithm of performance vs the logarithm of time. While Goddard provided a start at comparing these different methods of measuring technical progress, the next section will describe the different pros and cons of each method and how they relate to each other.

### 2.1.4. A comparison of the different methods of measuring technological improvement change

Each of the methods discussed in this chapter presents different theoretical and practical arguments for why technical progress should be presented in that manner. Out of these many methods of charting technological progress, the ones presented by Moore (1965) and Wright (1936) have shown to be the most robust (Nagy et al, 2013). In this section we describe the debate about which of the two most prevalent methods of measuring technical change should be used. The end of this section will discuss an important paper by Nagy et al (2013) who show empirically (following earlier theoretical work by Sahal) that there are few differences between Moore's Law and Wright's Law and that they can both be used effectively.

McDonald and Schrattenholzer (2001) criticize Moore's Law when they compare Moore's Law to the aging of a fine vintage wine and state that instead the accumulation of experience (or production) is what leads to technological cost reductions.

*For most products and services, however, it is not the passage of time that leads to cost reductions, but the accumulation of experience. Unlike a fine wine, a technology design that is left on the shelf does not become better the longer it sits unused. (McDonald and Schrattenholzer, 2001)*

The opposing side of the debate has been described by Nordhaus (2008) in his paper 'The Perils of the Learning Model for Modeling Endogenous Technological Change,' in which he states that:

*Learning has become a favorite tool for modeling technological change in many models of the energy sector and of global warming. It is convenient because learning-by-doing is one of the few "theories" of technological change that is easily included in models because of its simple specification. It is a dangerous modeling technique, however, because the learning rates are biased and because it therefore seriously underestimates the marginal cost of output. (Nordhaus, 2008)*

Nordhaus argues that there is a fundamental identification problem with the production-based models in that they cannot reliably separate production from exogenous (non-learning) technological improvement.

While this debate has received a good amount of attention, a recent paper by Nagy et al (2013) shows empirically that the theories may be more similar than most people think. In this paper they describe that production of goods tends to increase exponentially, and therefore

*'A combination of an exponential decrease in cost and an exponential increase in production would make Moore's law and Wright's law indistinguishable.' (Nagy et al, 2013)*

While the paper by Nagy et al (2013) is the most recent example of comparing the methods of measuring technical change, Sahal (1979) was the first to mathematically relate Moore's Law and Wright's law. Sahal (1979) describes the curves that relate technological progress to a dependent variable (time, production, etc.) as 'progress functions' and describes the exponent of the progress functions (what we refer to as 'R' above) as the critical variable. In the study, Sahal then relates the progress functions of Moore and Wright to each other. This relationships is summarized by Nagy et al (2013) as w = m/g, where w is the power law growth parameter in Wright's Law, m is the exponential growth parameter in Moore's Law and g is the exponential growth parameter for cumulative production. The various technical improvement curves can be directly converted between Wright's Law and Moore's Law by using the relationship between cumulative production and time. It is for this reason that throughout the rest of the paper, time will be used as the dependent variable when measuring technological change except for in section 4.4, where several different effort variables will be compared using the results of this thesis.

Beyond relating Moore's and Wright's frameworks, Sahal was one of the first to explicitly question what causes the variation in the exponential growth parameters.

*'It should be noted that the progress function has many variations in form and parameters. Nevertheless, the exponent of the function is common to certain cases...In particular, operations with similar ratios of assembly to machine work turn out to have similar progress function exponentials.' (Sahal, 1979)*

Sahal posits that the ratio of human to machine work time could be an explanatory variable for the improvement rates of technologies. While this particular theory is not explored further in Sahal's study or this thesis, the question of finding indicators that can help predict TIRs is at the core of this thesis.

The next chapter will explore further some of the theoretical explanations of technological progress that can lend credibility to the long-term usefulness and applicability of the improvement functions mentioned in this chapter.

## 2.2. Why Technologies Change Over Time

Section 2.1 demonstrated several ways to track technological change and provided evidence to show that technological domains exhibit exponential improvement over time. Section 2.2 builds upon the idea of improvement rates and describes several theories as to why technology may changes over time. In particular, this section explores theories that attempt to explain the underlying mechanisms that drive technological improvement. Coming to a complete understanding of all of the mechanisms that contribute to how technology improves over time is likely to be futile, as the problem contains a countless number of social and technical interactions. A goal of this research is to try to discover a selection of the strongest individual contributing factors to technological improvement, while acknowledging that these factors are certainly not exhaustive.

Many of the theories on why technologies improve over time are qualitative in nature and use data sources such as interviews, surveys, and observations in their case studies. In this thesis the theoretical underpinnings from each of these theories is synthesized into a set of hypotheses that are tested via patent based heuristics. In this way, the qualitative information that has been gathered in the field can be complemented by quantitative data that can be used to improve upon current theories and better inform future qualitative theories.

This process requires an understanding of the current literature on technological change and the resulting theories. It seems to the author that the most logical way of progressing through the extensive literature on technological change is chronologically, and thus this section will begin with Shumpeter's ideas on creative destruction and technological paradigms (1939) and end with modern day theories on why technologies change over time. While the review in this section will contain some of the most well known theories on technological change, these are by no means the only sources about technological change. A number of the lesser-known theories will be discussed in the final sub-section, but they will not be discussed in great depth for the sake of focus and brevity.

### 2.2.1. Schumpeter and creative destruction

Schumpeter was one of the earliest contributors to the field of technological change, and he made a clear delineation in the difference between the terms 'invention' and 'innovation.' He defines innovations as 'changes in production functions which cannot be decomposed into infinitesimal steps', and follows this with the example:

*'Add as many mail-coaches as you please, you will never get a railroad by so doing.'* (Schumpeter, 1935)

'Inventions', on the other hand, are experiments that do not in themselves exert any influence on business life at all' (Schumpeter, 1935). Building upon these seminal definitions, Schumpeter introduced the idea of creative destruction, in which he posits that economic progress is driven by firms providing goods and services who are then displaced by new firms who create improved goods and services, who, in turn, are displaced by the next generation and so on. He denotes this cyclical process of innovation and obsolescence 'creative destruction'. Caballero and Jaffe (1993) provide an excellent quote from Schumpeter about this theory:

*'The fundamental impulse that sets and keeps the capitalist engine in motion comes from the new consumers' goods, the new methods of production or transportation, the new markets, the new forms of industrial organization that capitalist enterprises creates .... [examples] ... [these examples] illustrate the same process of industrial mutation that incessantly revolutionizes the economic structure from within, incessantly destroying the old one, incessantly creating a new one. This process of Creative Destruction is the essential fact about capitalism . . . (Joseph Schumpeter, 1942)'*

The idea that innovation can drive the economy and technology forward has proven to stand the test of time, as his work is still cited frequently by many of the other authors cited in this section. Relating Schumpeter's work to the question of *why* technology changes over time, he describes the idea of older technologies being replaced by more recently invented ones that rely upon new ideas, and that this will happen again. Technology improved because new ideas and technologies are discovered that are significantly better than the technologies that are being replaced.

### 2.2.2. Vannevar Bush and the linear model

Much of this sub-section is derived from the excellent review of the linear model provided by Balconi et al (2010). To say that the linear model has been impactful in the field of technological development would be a massive understatement. Freeman (1996) claimed that at one point in time it was nearly impossible to read an article related to technological change or related policies without discussing the linear model. While some debate the origins of the linear model, many consider Vannevar Bush's (1945) paper as the first complete publishing of the linear model. In this paper, Bush states that

*'Advances in science when put to practical use mean more jobs, higher wages, shorter hours, more abundant crops, more leisure for recreation, for study, for learning how to live without the deadening drudgery which has been the burden of the common man for ages past... But to achieve these objectives ... the flow of new scientific knowledge must be both continuous and substantial' (Bush, 1945)*

The core ideas that were introduced in this paper helped forge the foundation of the NSF that was created in a large part by Vannevar Bush. While Bush introduced the idea that basic science paves the way for applied research and improvements in society, several others have added to his ideas to create the backbone of what is now considered the linear model and is shown in figure 7.


**Figure 7: Summary of Linear Model (Godin, 2006)**

Godin (2006) describes how Bush modified his connection between basic and applied research around 1960 to include the idea of development. Finally, economists from business schools extended the model to non-R&D activities such as diffusion of innovations. In the end the model is often used as a starting point for what *not* to do in innovation research as it is often oversimplified and overgeneralized, but it nonetheless remains a cornerstone of many technological improvement theories today. The linear model is the main proponent of the idea that technological change is caused in a large part by basic scientific research.

### 2.2.3. Solow - Labor and Capital Growth

Solow (1957) approaches the technical change question from an economist's point of view and attempts to estimate the technical improvement of the US as a whole by analyzing the GDP increases between 1909 and 1949:

*What I want to describe is an elementary way of segregating variations in output per head due to technical change from those due to changes in the availability of capital per head. (Solow, 1957)*

His model attempts to explain the growth in GDP $q$ by the increases in capital, $k$, labor, $I$, and in technical capability, A(t), as is shown in equation 5.

$$q = A(t) * f(k, I) \qquad \text{(Equation 5)}$$

Solow's measure of technical capability is a proportional and therefore unit-less index of technical capability that is derived from the other components of equation 5, and is pegged at 1.0 in 1909 and rises to 1.81 in 149.

The resulting time series of technical indices for the nation allowed Solow to distinguish between GDP increases due to capital, labor and technical improvement. First, Solow discovered that his technology index A(t) increased at approximately 1.5% per year over the course of the 40-year period. He later refines this when he describes the increase being ~1% between 1909 and 1929 and 2% between 1930 and 1949. This value of broad technological progress indicates a relatively consistent exponential improvement rate of all of technology and even hints at an increasing rate, which would show a hyper-exponential increase of the total technological capabilities of the nation.

Over the course of the 40-year study, Solow shows that gross output per man-hour doubled. He then calculates that approximately 87.5% of this increase is due to technical change (better tools) and only 12.5% is due to increased use of capital (more tools). This finding is striking and places even more importance on to why studying technological change is important, as it may account for nearly 90% of the increase in human productivity.

Ultimately Solow showed that while increasing labor and larger capital investments play a role in GDP growth, technical change is also a large contributor that should not be discounted. This is especially important when considering the idea that improvements in output come from economies of scale or from production increase alone. Solow is quick to note, however that technical improvement does not come without labor working on improving technologies (R&D) and capital investments to implement the new technologies (replacing old machines with new ones).

### 2.2.4. Arrow and learning by doing

Arrow (1962) introduced his idea of learning by doing as an attempt at helping to explain the improvement of technology over time. He begins his argument by claiming that while trends of technological improvement can be practical and useful, they are essentially ignoring causation in support of correlation.

*'From a quantitative, empirical point of view, we are left with time as an explanatory variable. Now trend projections, however necessary they may be in practice, are basically a confession of ignorance, and, what is worse from a practical viewpoint, are not policy variables.' (Arrow, 1962)*

Arrow refers to the work of Wright (1936) as well as several other studies that show that the product output of a technology (output per man-hour) can rise in the absence of specific investment in new tools or practices through non-formal learning channels. One excellent example is the case of the Horndal Iron works in Sweden, where they experienced productivity improvements of approximately 2% per year over the span of 15 years without any specific investments in R&D (Lundberg, 1961), although it is almost certain that they benefitted from the advancements from other companies within and outside of the field during this time period. It is cases such as these that lead Arrow to introduce his theory of Learning by doing.

Arrow's model can be summarized by the following oft-cited quotation from the paper.

*'technical change in general can be ascribed to experience, that it is the very activity of production which gives rise to problems for which favorable responses are selected over time' (Arrow, 1962)*

Arrow generalizes the idea of knowledge and says that all technical improvement comes from experience working with a technology, and uses the cumulative production of the capital goods as the main index of experience. Although he credits learning and producing goods as the driving factor in technological improvement, there is a statement in the limitations of the study that mentions that although his model attributes learning mostly to the production of goods, that it can also take place through research at universities and other institutions.

*'It has been assumed here that learning takes place only as a by-product of ordinary production. In fact, society has created institutions, education and research, whose purpose it is to enable learning to take place more rapidly. A fuller model would take account of these as additional variables.' (Arrow, 1962)*

Arrow's model has often been used as a theoretical backing to the production based technological graphing methods such as Wright's Law, but the prior statement clearly indicates that he leaves room for the effect of non-production based learning such as R&D in universities, national labs and private companies. Arrow's theories indicate that technologies improve over time through the accumulation of producing the technology and studying it in a specific R&D setting.

### 2.2.5. Dosi and market-pull vs technology-push

While many previous attempts at understanding how technology changes over time are looking for overarching theories to explain progress, Dosi (1982) describes the goal of his seminal paper as much more modest and potentially realistic.

*'This paper does not aspire to provide a "general theory" of technical change. It simply attempts to focus on questions like "Why did certain technological developments emerge instead of others?" "Are there regularities in the process of generation of new technologies and in technical progress thereafter? Is there any regularity m the functional relationship between the vast number of economic, social, institutional, scientific factors which are likely to influence the innovative process?" ' (Dosi, 1982)*

This is an important departure from prior works, as it began a trend of looking more closely at specific determinants of technological change in more focused studies. Dosi recognized that the factors that contribute to the progress of technology might be numerous, varied, and interrelated.

*The origin of the latter [technological progress] stems from the interplay between scientific advances, economic factors, institutional variables, and unsolved difficulties on established technological paths. (Dosi, 1982)*

Beyond this very important reframing of the problem, Dosi centered his research on one of the most popular theories of what drives technological progress: 'Demand (or Market) Pull' vs 'Technology Push'. The 'demand-pull' side of this debate centers around the idea that technology is only developed based upon what the market needs, and thus progress is driven by the markets 'pulling' technology forward. On the contrary, the technology push arguments states that as technology develops, new products are capable of being produced, and thus these new or less expensive products fill a demand. Dosi took an intermediate stance in this discussion and claimed that the process is more complex than one side or the other.

*'One-directional explanations of the innovative process, and in particular those assuming "the market" as the prime mover, are inadequate to explain the emergence of new technological paradigms.' (Dosi, 1982)*

Dosi goes on to point out particular flaws in both the demand-pull and technology push theories.

*'We will try to show that these latter interpretations [demand-pull] present a rather crude conception of technical change, as an essentially reactive mechanism, based on a "black box" of readily available technological possibilities. Moreover this conception contradicts substantial pieces of empirical evidence. On the other hand, extreme forms of technology-push approaches, allowing for a one-way causal determination (from science to technology to the economy) fail to take into account the intuitive importance of economic factors in shaping the direction of technical change.'*

*(Dosi, 1982)*

This rebuttal of both sides of the market vs technology debate allowed the debate to widen into more complex understandings of how technology improves over time. Dosi contributes the idea that the demand for a technology combined with its technological potential are key factors that determine the improvement of a particular technology.

### 2.2.6. Christensen's Theory of Disruptive Innovations

One of the more recent theories of how technologies improve over time is from Clay Christensen and his theories on disruptive innovations. Christensen (1997) describes the 'Innovator's Dilemma' in which new inventions and innovations begin with capabilities that are below that of the existing products along the main metric of choice but serve a niche market and prove to have a higher capability in another, less regarded capability. An example of this is the ARM processor that is used in many smartphones. The ARM processor is generally regarded to be not as fast as the traditional x86 processors manufactured by market leader Intel, but they are more energy efficient. Prior to the introduction of smartphones, the energy efficiency of computer processors was far less important than the overall speed of the chip. This changed drastically when smartphones were introduced and more energy efficient processors were needed for the pocket-sized devices to last all day.

Christensen provides other examples of how a technological domain can be replaced by another that is improving more rapidly when measuring a different metric. His most famous example is that of disk drives used in computer information storage.

*'For example... shipments of 5.25-inch products in the 30-100 mb range in the total market first surpassed unit shipments of 14- and 8-inch drives in 1984, when areal density of the new architecture was still nearly 40% below that of 8-inch products. In the next generation, 3.5~inch unit volume surpassed all earlier architectures in the 30- 100 mb category in 1988 and in the 100-300 mb category in 1989, even though their densities were still substantially inferior to those achieved in the prior architectures. This is because 5.25~inch products were first used in desktop computing and 3.5~inch drives in portable computing, where the metrics of performance were very different than the simple area density measure that had been sufficient when evaluating larger drives used with larger computers.'* (Christensen, 1992b)

He states that the improvement rates of the 5.25" disk drives are higher than that of the 8- and 14-inch drives that were replaced. This cycle happened again when the 3.5" drives displaced many of the 5.25" drives, this time due to the improved performance in a different functional performance metric (FPM) - most likely a measure of efficiency related to reduced power consumption that is more important in mobile computing than in tradition desktops.

While Christensen provides many specific case studies about the improvement of technologies, he also created several theories to back them up. Christensen's main hypothesis concerning why technologies improve over time is that new technologies displace old technologies by adopting technologies that improve more rapidly in new and different functional performance metrics that were focused on previously.

Christensen also tends to rely heavily upon the idea of technological S-curves, and uses them to graphically show how technologies improve. Figure 8 shows how a technology (in this case "B") can seem inferior to another technology "A" when measuring it using one FPM "as defined in Application A", but can seem superior when considering it on another FPM "as defined in Application B".

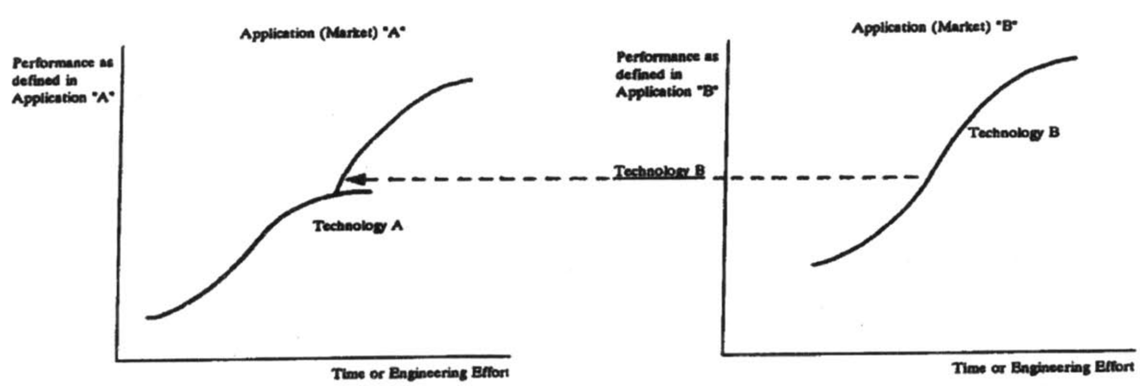



**Figure 8: Christensen's model of technological disruption depicted using competing S-Curves (adapted from Christensen, 1992b)**

Relating this figure to the example of hard-disks, Technology A could represent 5.25" hard disks, 3.5" disks are technology B, with the desktop and mobile computing markets as applications "A" and "B" respectively. The 'S-Curves' sub-section will further explore the use of s-curves in the technological change literature.

#### 2.2.6.1. Properties of S-Curves

Christensen's decision to use S-curves as the foundation for his theory of disruptive innovation makes him one of many technological change researchers who subscribe to this nearly ubiquitous theory of how technologies change over time. Sood and Tellis (2005) claim that 'belief in this [S-Curve] premise is so strong that it has become almost a law in the strategy literature,' however they note that there has still not been 'any single, strong, and unified theory for the S curve,' and that 'there is scattered empirical support for the premise and limited theoretical support for various aspects of the S-shape curve.' For the sake of discussion, Sood and Tellis (2005) created a rough summary of the S-Curve theories:

A central premise is that performance of a new technology starts below that of an existing technology, crosses the performance of the older technology once, and ends at a higher plateau, thus tracing a single S-shaped curve (see Figure 9).

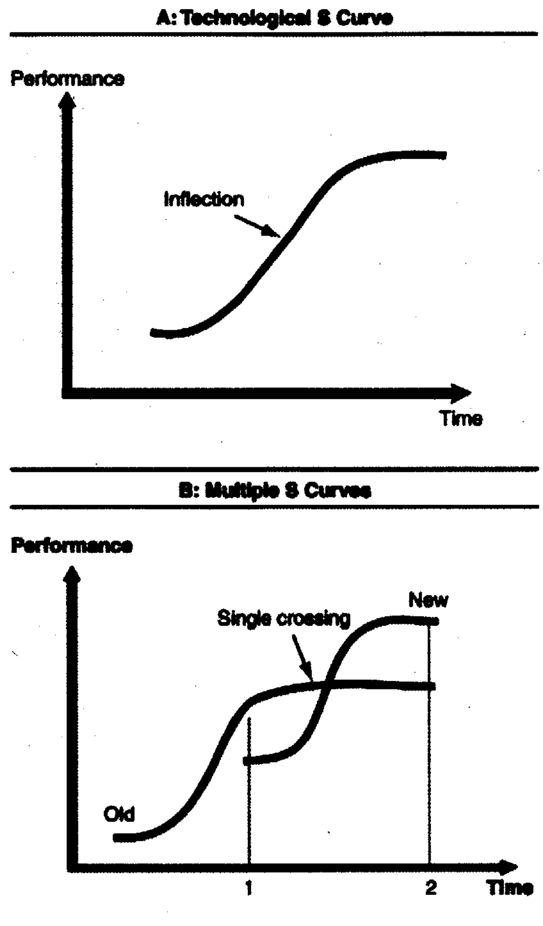



**Figure 9): Demonstration of the basic theory of S-Curves, adapted from (Sood and Tellis, 2005)**

Sood and Tellis claim that there are 3 main stages that make up the logic of the S-Curve: introduction stage, growth stage, and maturity stage. The introduction stage is slow moving for a new technology due in part to the fact that it is not well known and thus does not have significant attention of researchers or practitioners. The growth stage then emerges when a dominant standard or architecture is decided upon and more significant resources are devoted to the development of the technology (Utterback, 1974). Finally, the maturity phase is reached where the technology levels off and the improvement rate of a technology is much slower than in the growth stage. Many reasons for the plateauing of improvement rates have been hypothesized over the years. Foster (1986) claims that there is a natural limit for each technology and that there can only be a certain amount of improvement to be had before the limit is reached. Sahal (1981) cites scaling limits as the reason for this decline - citing that when technologies reach very large or very small scales, they become very difficult to improve upon. Regardless of the specific conjecture, the maturity stage is often regarded as the 'limit' of the technology, and occurs when there becomes an engineering bottleneck that is either too difficult or too expensive to solve.

#### 2.2.6.2. Empirical Rejection of Technological Limits

While the technological literature seems to be universally accepting the idea of S-curve limits, there have been some notable detractors to this idea including Christensen.

'An explanation of why Fujitsu and CDC perceived limits to be at such different levels is that <u>nobody knows what the natural, physical performance limit is in complex engineered products</u>, such as disk drives and their components. Since engineers do not know what they may discover or develop in the future, since the <u>physical laws (and the relationships between laws) governing performance are imperfectly understood, and since possibilities for circumventing known physical limits cannot be well foreseen</u>, the natural or physical limits cited by scholars of technological maturity, such as Foster (1986) and Twiss (1979), may <u>in practice be moving targets rather than immovable barriers</u>. (Christensen, 1992b)

It is evident in this quote by Christensen that he is not entirely certain in the assumption of definite technological limits, but rather technological plateaus that simply take longer to overcome than their other counterparts. One excellent example of an industry overcoming its 'limits' was presented by Henderson (1993) in her review of the photolithography equipment industry. In the paper, she states that 'Unexpected changes in user needs and in the capabilities of component and complementary technologies permitted optical photolithography to dramatically exceed its 'natural' limits.' Figure 10 shows how the realized performance of the resolution of optical photolithography compared with the ever-changing estimates of predicted limits, and especially illustrative is how the performance has exceeded the 'limits' of past predictions.

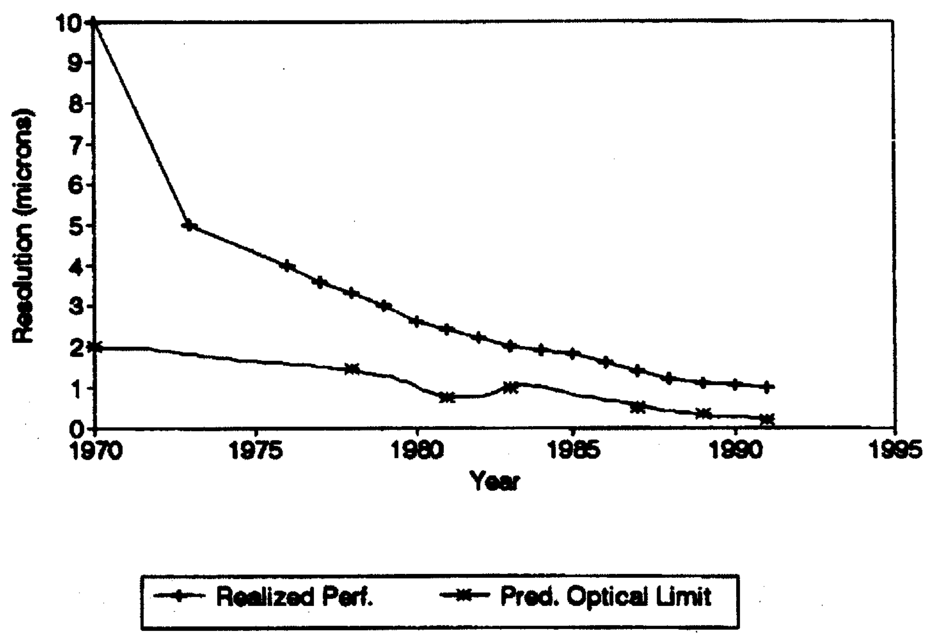



**Figure 10: Photolithography performance compared with predicted limits - adapted from (Henderson, 1995)**

The prior examples bring into question the validity of the 'maturity' stage of the S-Curve theory.

Another area where the S-curve has been questioned is in the general scale of the measurement of performance and what is the appropriate independent variable (horizontal axis). As was mentioned in previous sections in this thesis, there are often discussions in technological literature over the correct independent variable for the measurement of technological change. The two main theories involve comparing improvements with time and effort respectively.

In Christensen's (1992A) thorough examination of the improvement rates of the computer disk drives, he provides an excellent comparison between the improvement in areal density of the hard-disks with time and with cumulative industry revenues as a proxy for engineering effort, and is shown in Figure 11

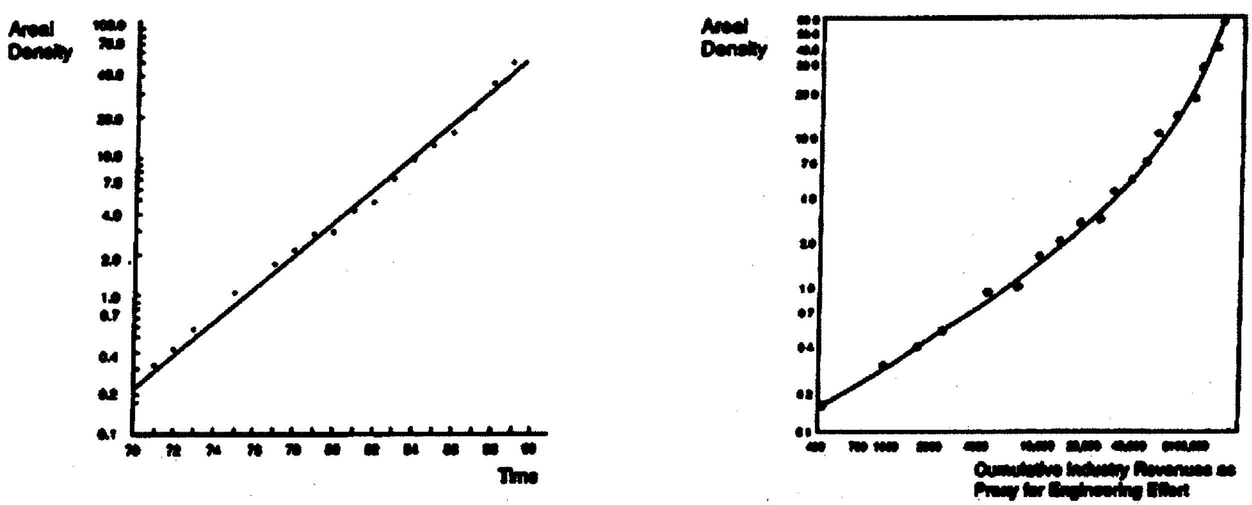



**Figure 11: Christensen's comparison of the log of areal density of disk-drives vs time (left) and log of areal density vs log of industry revenues (as a proxy of engineering effort) - adapted from Christensen (1992A)**

The conclusion that Christensen draws from this comparison is that 'a relatively constant rate of improvement over time in areal density appears instead to be an increasing rate of improvement per unit of engineering effort applied.' Which appears to be correct at first glance, but the chart of performance vs time is charted on a Log-Linear graph and the performance vs revenue is charted on a Log-Log graph. If the effort graph is re-drawn on log-linear scale, then the comparison looks much different, with a sharp increase in performance when revenues are low and a leveling out to a nearly constant rate of improvement with higher revenues in the industry, showing a very similar look to that of the performance vs time graph. If the performance vs effort graph is redrawn on a linear-linear scale, it looks very similar in shape to a pure exponential that is represented by the performance vs time graph. The redrawn graphs are shown in Figure 12.

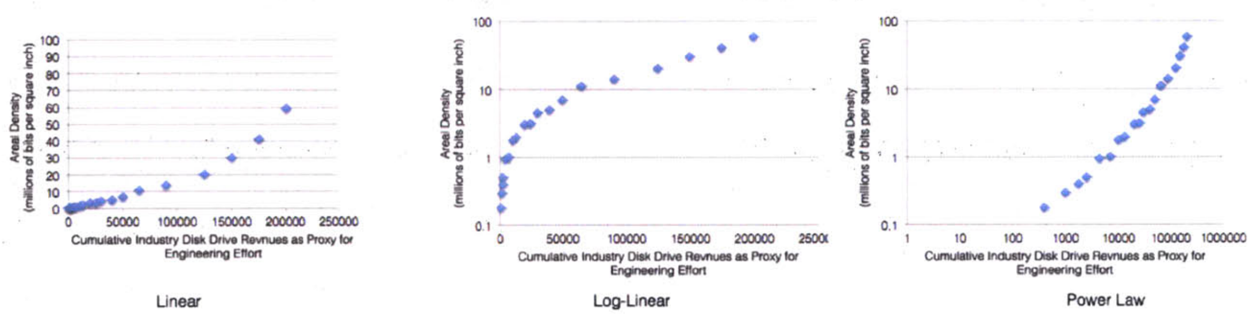



**Figure 12: Plots of disk-drive areal density vs industry revenues for 3 different log scales - adapted from Christensen (1992A)**

Ultimately the lack of specificity of the scales of the axes in performance curves (of which S-Curves are a subset) can lead to significantly different conclusions for the same set of data.

One of the final reasons why the theory of S-curves has been questioned is the relatively small amount of quantitative data to support it. Continuing on in Christensen's study of disk-drives, he publishes a set of 'S-Curves' for the improvement of areal density vs time for 2 different companies in the 1970s and 1980s as shown in Figure 13.

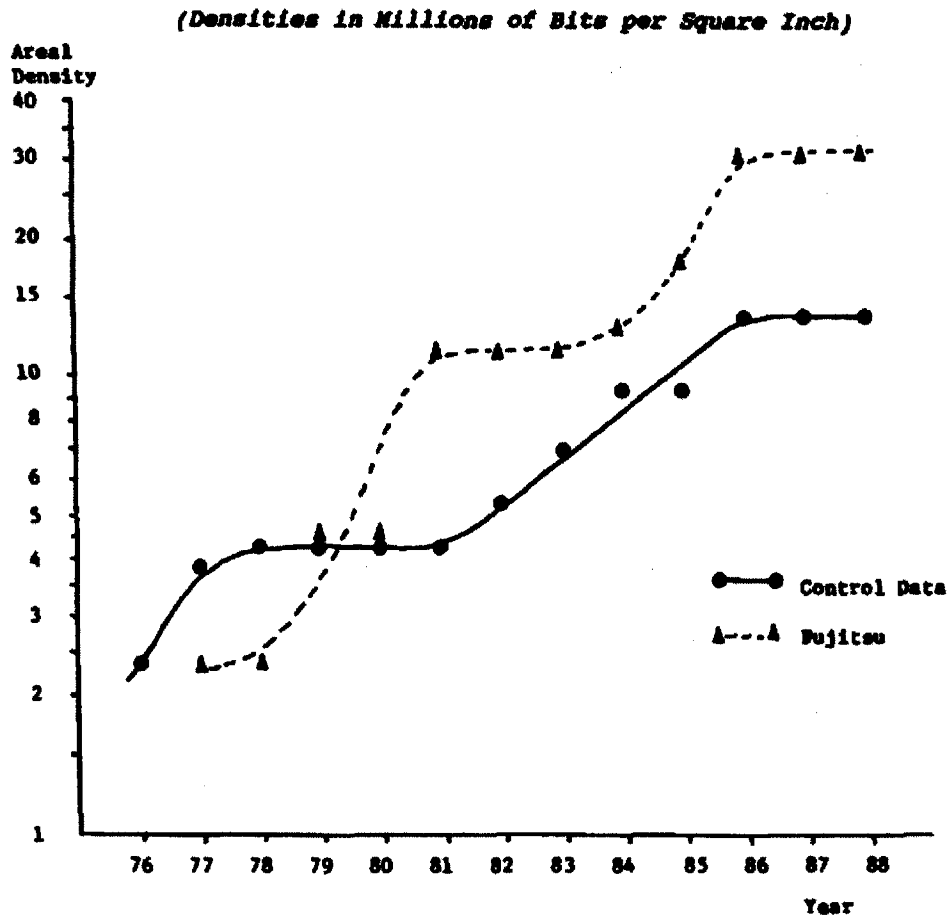



**Figure 13: Improvement curves for 2 companies of logarithm of disk drive areal density vs time - adapted from Christensen (1992A)**

In his figure, Christensen shows a set of points that are connected by curves that resemble multiple s-curves linked together. Another way of looking at this data would be to fit it to an exponential regression, and see if it displays the constant rate of improvement (constant slope on a log-linear graph) that many other technologies demonstrate. Figure 14 shows each of those curves fitted to an exponential along with the combination of the two fitted to one exponential combined.

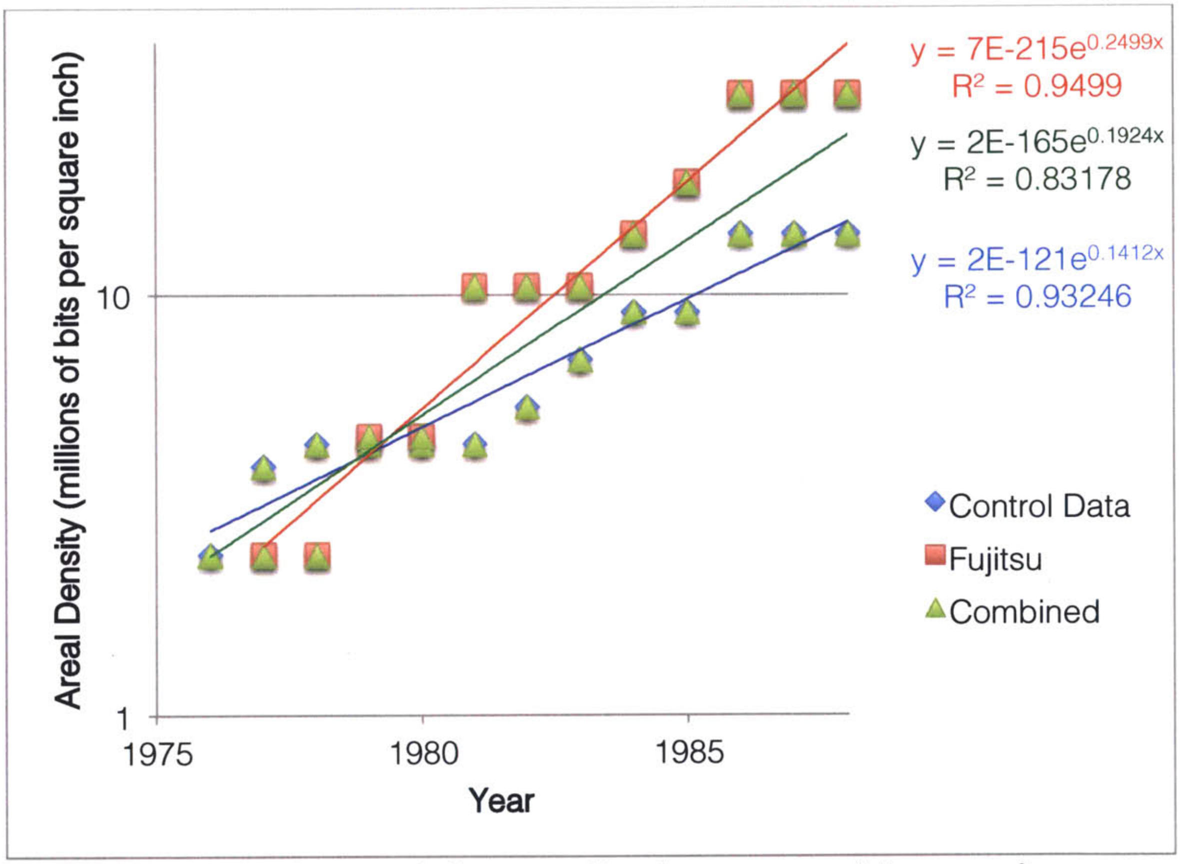



**Figure 14: Christensen's S-curves fitted to exponential regressions - adapted from Christensen (1992A)**

When the apparent s-curve data is fitted to an exponential, the fits are very strong, with an R2 of 0.95 for the Fujitsu data, 0.93 for the Control Data Corporation data and even a relatively strong 0.83 for the combined data. These strong fits show that while it is possible to loosely connect the data points into s-curves, the long term relationship between these data points is consistent with a constant exponential.

It is reasonable at this point to wonder why so the s-curve theory is very popular when the evidence supporting it is not particularly strong. One possible answer to this is that the s-curve theory is simply an adaption of a diffusion curve with limited resources. S-curves curves are common in traditional diffusion literature and have been shown to be quite accurate for the diffusion of many different technologies (Comin and Mestieri, 2013). The s-curve's theoretical support in diffusion is stronger as well, as there is often a limit built into the definition, for example the percentage of a population that owns a cell phone can never go over 100%, and thus the maturity stage is much easier to accept in the diffusion literature than in that of technological improvement.

### 2.2.7. Classifying Inventions

One of the main strategies used by many technological change researchers is to explain why technologies change over time by categorizing the improvements in different ways. Essentially, there exists a large cohort of researchers who claim that technology improves in leaps and bounds due to very important inventions. There have been many names for variations on this theory, consisting of terms such as: revolutionary, breakthrough, discontinuous, process innovation, radical, architectural innovations, dominant design, or disruptive. Many of these terms are 'intrinsically problematic because they define an innovation in terms of its effects rather than its attributes,'(Sood and Tellis, 2005).

One example of attempting to categorize specific improvements in a technological field is Tushman and Anderson's 1986 paper on technological discontinuities. They claim to demonstrate that 'technology evolves through periods of incremental change punctuated by technological breakthroughs,' and that 'those firms that initiate major technological changes grow more rapidly than other firms.' The underlying message of this theory is that technological change is faster and more impressive with a higher number of very important inventions.

In almost all cases of innovation categorization theory, there is both a lesser and a greater classification. For example, incremental change is slow and bit-by-bit, while radical change can happen all at once. Similar statements can be made for component vs architecture, and incremental vs breakthrough. Each of these qualitative description attempts to define how to categorize an innovation, of which Tushman and Anderson provide great examples:

Major technological innovations represent technical advance so significant that no increase in scale, efficiency, or design can make older technologies competitive with the new technology (Mensch, 1979; Sahal, 1981). Product discontinuities are reflected in the emergence of new product classes (e.g., airlines, automobiles, plain-paper copiers), in product substitution (e.g., transistors vs. vacuum tubes; diesel vs. steam locomotives), or in fundamental product improvements (e.g., jets vs. turbojets; LSI vs. VSLI semiconductor technology) (Tushman and Anderson, 1986)

While the definitions of the greater or lower classification are often detailed, they are also almost always subjective and open to interpretation. This means that oftentimes the decision of whether an invention is upper or lower class can be different based upon the researcher, which reduces the repeatability of the theories derived from these subjective determinations. For example, in their review of breakthrough inventions, Tushman and Anderson described the process of selecting their innovations as easy, but have very little detail regarding their selection process beyond that. Technological discontinuities were relatively easy to identify because a few innovations so markedly advanced the state of the art that they clearly stand out from less dramatic improvements (Tushman and Anderson, 1986)

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

**Significant Technological Discontinuities**

| Industry | Year | Event | Importance | Type of discontinuity | Locus of Innovation New firms | Existing firms | Probability |
|---|---|---|---|---|---|---|---|
| Cement | 1872 | First production of Portland cement in the United States. | Discovery of proper raw materials and importation of knowledge opens new industry. | Niche opening | 10 of 10 | 1 of 10 | |
| | 1896 | Patent for process burning powdered coal as fuel. | Permits economical use of efficient rotary kilns. | Competence-destroying | 4 of 5 | 1 of 5 | .333 |
| | 1909 | Edison patents long kiln (150 ft.). | Higher output with less cost. | Competence-enhancing | 1 of 6 | 5 of 6 | .001* |
| | 1966 | Dundee Cement installs huge kiln, far larger than any previous. | Use of process control permits operation of very efficient kilns. | Competence-enhancing | 1 of 8 | 7 of 8 | .000* |
| Airlines | 1924 | First airline. | Mail contracts make transport feasible. | Niche opening | 9 of 10 | 1 of 10 | |
| | 1936 | DC3 airplane. | First large and fast enough to carry passengers economically. | Competence-enhancing | 0 of 4 | 4 of 4 | .005* |
| | 1959 | First jet airplane in commercial use. | Speed changes economics of flying. | Competence-enhancing | 0 of 4 | 4 of 4 | .005* |
| | 1969 | Widebody jets debut. | Much greater capacity and efficiency. | Competence-enhancing | 0 of 4 | 4 of 4 | .005* |
| Minicomputer manufacture | 1956 | Burroughs E-101. | First computer under $50,000. | Niche opening | 1 of 8 | 7 of 8 | |
| | 1965 | Digital Equipment Corp. PDP-8. | First integrated-circuit minicomputer. | Competence-destroying | 3 of 6 | 3 of 6 | .019* |
| | 1971 | Data General Supernova SC. | Semiconductor memory much faster than core | Competence-enhancing | 0 of 7 | 7 of 7 | .533 |

**Table 1: List of Technological Discontinuities for three fields - adapted from Tushman and Anderson (1986)**

When looking at the table, there are a wide variety of inventions that are classified as breakthrough, including the first production of commercial cement and the introduction of a longer (150 ft) kiln for producing cement. It is possible that while these inventions received a significant amount of attention, that they were enabled by other inventions that may have proven to be just as important, yet less well known. This causes issues because for every Watt steam engine that gets the majority of the credit, there is a Wilkinson boring machine that enabled the engine to have precise and concentric cylinders, for every transistor there is a point rectifier for a radio that demonstrated the initial principle first. The purpose of these examples is to show that while we may remember one specific invention as being the most important, it is often one of many inventions that together were able to create a new and successful product or product class.

Table 2 shows another example of an attempt to classify significant innovations throughout history, with a list of important innovations throughout the last 300 years as given by Girifalco (1991).

| Innovation | Year |
|---|---|
| Seed Drill | 1731 |
| Watt Engine | 1776 |
| Rifle | 1824 |
| DC Motor | 1872 |
| Airplane | 1903 |
| Methyl Methacrylate | 1935 |
| Transistor | 1947 |
| Personal Computer | 1974 |

**Table 2: Example list of innovations throughout history (Girifalco, 1991)**

From examining this list, it is clear that there are issues in deciding what constitutes an important invention. These issues stem partially from ambiguity about the *level* of innovation to be considered. For example, the transistor is shown in the list and is widely considered a breakthrough technology, at the same time so is the integrated circuit (an important way of utilizing and manufacturing transistors) and so has the personal computer (also shown in the list and which depends directly on integrated circuits). It is also reasonable to consider the entire field of information technology (which relies on personal computers and many other technologies) the most important technological breakthrough of the latter half of the 20$^{th}$ century.

The ambiguity is further compounded by the fact that the described technological improvements can be collectively combined over varying time periods. In fact, this is often done to simplify communication about developments within a field. For example, the initial invention of the transistor was completed in a much shorter time than all of the ensuing and continuing changes in transistors. The same is true relative to the initial invention of the integrated circuit and the modern computer, which is only one aspect of the development of information technology.

It is interesting to note that many design changes designated as significant improvements or breakthroughs appear to identify individual points of improvement. Thus, the list in Table 1 would integrate well with the widely accepted conjecture that technological breakthroughs drive the overall improvement of a technology. The importance of breakthroughs is supported by Kaplan (1999), who states that

*'substantial growth over the long horizon requires discontinuous innovation'* (Kaplan, 1999).

A similar point of view is given by Ahuja and Lampert (2001) as they state that breakthrough inventions

*'serve as the basis of new technological trajectories and paradigms and are an important part of the process of creative destruction in which extant techniques and approaches are replaced by new technologies and products'.*

The idea that a small set of technological changes account for a large portion of the overall improvement has been echoed many times in regards to different aspects of technology.

It is for these reasons that their exists a need for a more quantitative and repeatable methodology of selecting important (or un-important) inventions and evaluating their impact on a specific technological domain. The following section will discuss how patents can be used as a proxy for inventions and can be objectively and repeatably categorized and analyzed.

## 2.3. Patents as a Proxy for Inventions

Throughout the rest of this thesis, patents will be a major source of data; this section will present a broad spectrum of literature on the use of patents as a proxy for inventions.

The chapter will begin with a short review of general information on patenting, including the rules for patenting, information included in patents, and the difference between the United States and International patenting practices. Next, seminal literature in the patent analysis field will be presented and discussed. Finally, limitations of the use of patents for technological analysis will be presented.

### 2.3.1. Overview of Patenting in the U.S.

This section of the chapter will give an overview of the patenting system. It will begin with a description of the US patenting process and the United States Patent and Trademark Office (USPTO). The types of information found in patents will be discussed next, ranging from the metadata to the structured text found within each document. Finally, the US and International Patent systems will be compared.

A United States patent gives its holder 'the right to exclude others from making, using, offering for sale or selling' the invention. In theory, this protection allows for the distribution and dissemination of the knowledge contained in the invention in exchange for a monopoly over the right to produce that particular invention for a limited period of time up to 20 years. There are several different types of patents that protect different types of inventions and have different uses and range from utility patents to patents on plants. In this thesis we are most concerned about utility patents, which are 'Issued for the invention of a new and useful process, machine, manufacture, or composition of matter, or a new and useful improvement thereof' (USPTO.gov, 2014).

An invention must fulfill three requirements in order to be patented, it must be novel, non-obvious, and useful. An invention is NOT considered novel if '(a) the invention was known or used by others in this country, or patented or described in a printed publication in this or a foreign country, before the invention thereof by the applicant for patent, or (b) the invention was patented or described in a printed publication in this or a foreign country or in public use or on sale in this country more than one year prior to the application for patent in the United States . . .' (USPTO.gov, 2014). Once the invention is considered novel, the differences between the invention and any current knowledge must be considered non-obvious by a person 'having ordinary skill in the area of technology related to the invention' (USPTO.gov, 2014). An example of an obvious invention is the substitution of one color for another or a change in size of a previous invention, neither of which is patentable. The USPTO refers to the term useful as meaning 'the condition that the subject matter has a useful purpose and also includes operativeness, that is, a machine which will not operate to perform the intended purpose would not be called useful, and therefore would not be granted a patent' (USPTO.gov, 2014).

Once granted, a patent typically grants the inventor 20 years of monopoly over that particular invention. Some patents may not last that long, as the USPTO requires maintenance fees at shorter intervals along the way partly to ensure that inventions are being made use of while the monopoly is in effect. Additionally, some cases allow for the extension of a patent's monopoly for longer than 20 years (USPTO.gov, 2014).

### 2.3.2. Information contained in patents

Filing a patent can be an arduous process, as a considerable amount of information is included in the application and the final granted patent. This section will discuss some of the information that is contained within a patent. In order to do so, we will walk through reading an example patent.

Figure 15 shows the front page of a US Utility patent. This page typically contains most of the metadata that is associated with a patent. The patent number is listed in the upper right hand corner and is denoted in this patent by the reference [11] (as marked on Figure 15). The patents numbers (PN) are issued in the order that they are granted, therefore the first patent issued was PN 0000001, and this patent, which was granted in 1980, is PN 4234352, meaning that there were over 4 million patents issued between the founding of the USPTO and November 18th, 1980, which is the date that the patent was issued and is denoted in the diagram by reference point [45]. The title of the patent is reference [54], and the inventor and assignee are self-evident.

The next two fields [51] and [52] are the patent classification codes. These codes are determined by the patent examiners and are a rather complicated duopoly of hierarchies that will be explained in greater detail in Section 3.4.5when the patent searching methodology is discussed. The most important information about the patent classifications is that they are meant to represent particular technical domains (ex: 136 – Batteries: photoelectric and thermoelectric).

United States patent examiners assign classification codes within both the US patent classification and the international patent classification system for each US patent issued. There are some differences between the two classification systems as is described by Gruber et al (2013).

*'Technological classifications employing the IPC system are typically based on the information contained in the description of the technological invention as well as the examples, drawings, and claims provided in the application document. This is a key difference between the IPC system and the U.S. Patent Office Classification (USPOC). Whereas the USPOC classifies patents according to the claims stated within the application document (i.e., the scope of protection), the IPC system considers the complete technological information contained in the application document, and thus classifies patents with respect to the technologies associated with the invention (OECD 1994).' (Gruber et al, 2013)*

**United States Patent** [19] [11] **4,234,352**

**Swanson** [45] **Nov. 18, 1980**

[54] **THERMOPHOTOVOLTAIC CONVERTER AND CELL FOR USE THEREIN**

[75] Inventor: **Richard M. Swanson**, Los Altos, Calif.

[73] Assignees: **Electric Power Research Institute, Inc.**, Palo Alto; **Stanford University**, Stanford, both of Calif.

[21] Appl. No.: **928,103**

[22] Filed: **Jul. 26, 1978**

[51] **Int. Cl.**[3] .............................. **H01L 31/06**

[52] **U.S. Cl.** .............................. **136/253**; 136/256; 136/255

[58] **Field of Search** .......... 136/89 RT, 89 SJ, 89 CC

[56] **References Cited**

U.S. PATENT DOCUMENTS

| | | | |
|---|---|---|---|
| 3,331,707 | 7/1967 | Werth | 136/89 |
| 3,549,960 | 12/1970 | Wedlock | 357/30 |
| 3,929,510 | 12/1975 | Kittl | 136/206 |

OTHER PUBLICATIONS

B. D. Wedlock, "Thermo-Photo-Voltaic Energy Conversion," *Proceedings IEEE*, May 1963, pp. 694-698.
C. W. Kim et al., "A P-I-N Thermophotovoltaic Diode," *IEEE Trans. Electron Devices*, vol. ED-16, pp. 657-663 (1969).
General Motors Corp., "Study of a Thermophotovoltaic Converter," Final Report, Contract No. DA 36-03-9-AMC-02255(E), Feb. 1964.
R. J. Schwartz et al., "Thermophotovoltaic Cells," Final Report, Contract No. DAAB 07-72-C-0281 (Jun. 1974).

*Primary Examiner*—Aaron Weisstuch
*Attorney, Agent, or Firm*—Flehr, Hohbach, Test

[57] **ABSTRACT**

Disclosed is a thermophotovoltaic converter which includes a parabolic cone radiation concentrator portion and a processor portion including a radiator which absorbs concentrated radiation and generates incandescent radiation. A photovoltaic cell in close proximity to the radiator receives the incandescent radiation and generates an electrical voltage. The cell includes an intrinsic or lightly doped silicon substrate having a top surface for receiving radiation and a bottom surface having a plurality of diffused N and P conductivity regions arranged in rows. A titanium-silver layer overlays the bottom surface and conductively interconnects regions of one conductivity type and provides a reflective surface to photons which pass through the substrate.

**17 Claims, 10 Drawing Figures**

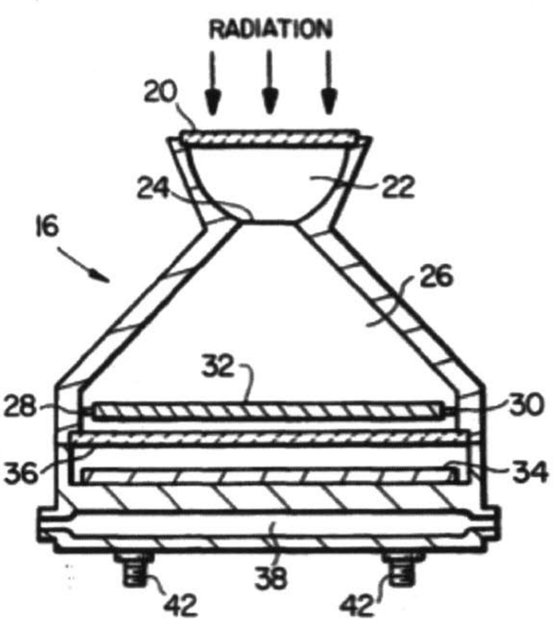



**Figure 15: Front Page of US Utility Patent showing UPC [51], IPC {52], citations [56], Non-Patent Literature Citations [OTHER PUBLICATIONS]**

It is important to note that each patent can have more than one classification in each system. In the example in Figure 15, the patent has one international classification code (H01L 31/06) and three US classifications (136/253, 136/256, 136/255). The information stored in these codes will prove to be an important point throughout this thesis, and thus more attention will be given to the topic in Chapter 3.4.

The field of search [58] shows the fields in which the patent examiner and the inventor searched to find prior art and/or references for the patent. The references that the patent makes to other patents are listed below and also form a pivotal source of information for this thesis. In addition to references to other patent documents, there is a list of 'Other Publications' that are references to non-patent literature, which in many cases are academic journal articles. The names of the patent examiners are followed by the abstract of the patent and then the list of figures.

Like the metadata, the text of each patent is structured and can be searched easily. The structure of the example patent is as follows: Abstract, Background of the Invention, Summary of the Invention, Brief Description of the Drawings, Description of the Preferred Embodiment, Claims. The background of the invention contains information on prior art and the problem that the invention is solving. The next three sections give a brief overview of what is novel about the invention and discloses some of the details. The most important legal part of the patent is the claims section, because the claims are what ultimately decide the breadth and the validity of the patent if it is challenged or discussed in court.

In addition to the information contained explicitly within the patent, there is also a significant amount of information that can be extracted from the organization of the patent system. In particular, patent citations are of great interest for this research. As is shown in Figure 16 each patent can cite other patents (backward citations) and can be cited by other patents (forward citations).

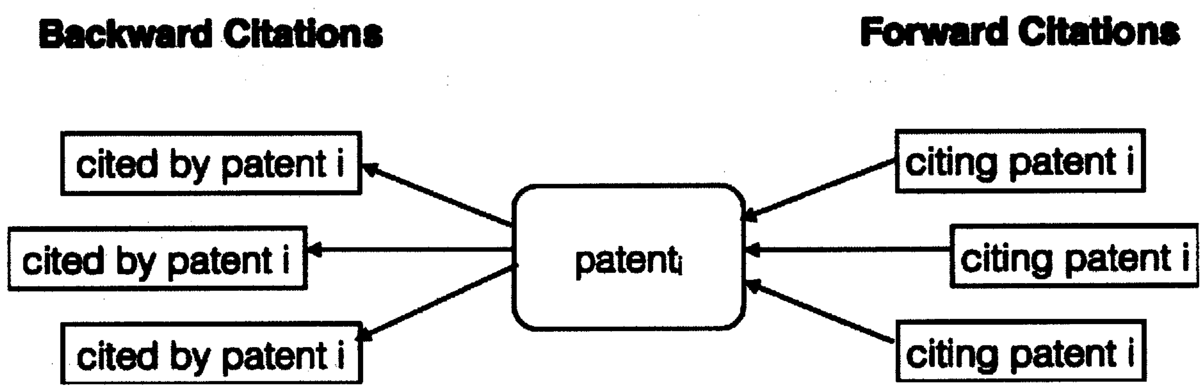



**Figure 16: Forward and Backward patent citations for patent *i* - adapted from Nemet and Johnson (2012)**

The citation network of patents allows for the defining of the relationship between different patents and is used extensively throughout this research.

### 2.3.3. Use of patents in technological research

One of the sources of data that has been widely used for understanding technological change in recent years is patent data. Patents are an attractive choice for analyzing technological change because they are: generalizable, objective, quantitative and qualitative.

Patents include many technical fields over a long period of time, and thus allow for easier generalization of the research. There are specific criteria for an invention to be patented, which creates an objective standard as to what counts as an invention. Each patent is well tracked and includes a wealth of meta-data, and thus allows for quantitative analysis. Additionally, each patent includes a significant amount of qualitative data to support and complement the quantitative analysis.

Campbell (1983) states that the patent database approximately records most of the advances in technology, which enables researchers to be confident in the completeness of their analysis. The use of patent data for economic and scientific analysis began in 1966 (Trajtenberg, 1990) and the growing capabilities of computers and data analytics tools have created a significant increase in the ability to search the patent data and extract useful information and insights in recent years (Joho et al, 2010). The amount of information that is easily accessible through the patent database and a web browser is orders of magnitude higher than what was available just 20 years ago (Michel and Bettels, 2001).

In addition to accessibility, there are other significant reasons why the patent database provides an excellent data source for analyzing technological change over time (Hall and Jaffe, 2001). Overall, patents are a set of data that contains the raw information created by the inventors of millions of patents over hundreds of years, and additionally the underlying information present in the organization of this massive data set that has been created by thousands of expert patent examiners. The combination of the data and organization potentially comprise the 'most valuable data in the world' (Atkinson, 2008). More effective use of this powerful patent information for understanding how technology grows over time is enabled if one can develop a robust, repeatable method for finding a relevant and complete set of patents and developing appropriate analytical tools to evaluate the sets of patents. The following sub-sections will explore prior attempts to search and classify patents, as well as some of the ways that patents have been used by researchers to develop and evaluate theories of technological change as was called for by well-known technological change researcher Manuel Trajtenberg.

'Once their meaning has been well established, the use of patent data may offer additional advantages in itself and over alternative data sources. First, patent data can be easily obtained all the way to the very beginning of a product class, whereas the gathering of conventional industry data usually starts only when a sector is well established. Thus, patent counts and citations may play an important role in studying the very emergence of new products, which seems to be the period when most of the important innovations occur. Second, patent data are richer, finer, and have a wider coverage than say, R&D expenditures, and are practically continuous in time.' (Trajtenberg, 1990)

#### 2.3.3.1. Patent Searching

The most basic ways of searching for patents are the keyword search and the classification search. The keyword search uses search terms and Boolean operators (AND, OR, NOT, NEAR) to construct queries to find the most relevant patents (Larkey, 1999). The classification search method requires that the patents already be classified (such as in the US or International Patent classification systems), and that the patent(s) in question can be pinpointed to just one or more patent classes (Baillie, 2002). Beyond the two most basic methods for retrieving sets from the patent database, there have been an increasing number of approaches involving complex information retrieval techniques and methods (D'Hondt, 2009). Table 3 shows a list of different approaches that have been used by patent researchers in recent years, which altogether makeup a patent searching 'toolbox'.

| Patent Searching Technique | Reference |
|---|---|
| Boolean | Baillie 2002 |
| US Patent Classification (UPC) | Baillie 2002 |
| International Patent Classification (IPC) | Takaki et al, 2004 |
| Query Expansion | Wang, 2011 |
| Dividing into different Time Periods | Wang, 2011 |
| Probabilistic Retrieval Models | Fujita, 2004 |
| Citation Linking | Fujii, 2007; Lopez and Romary, 2010 |
| Unigram and bigram frequency analysis | Magdy and Jones, 2010 |
| Knowledge representations of data | Graf et al, 2010 |
| Using Sample patent to generate keywords | Xue and Croft, 2009 |
| Semantic Analysis | Gerken and Moehrle, 2012 |

**Table 3: Patent searching techniques modified from Madhabi et al (2011)**

The techniques in Table 3 are a set of methods that can be combined in different ways to locate a specific set of patents, as demonstrated by Wang (2011). The methodology described and tested in this paper is a novel combination of the two simplest approaches in this listing.

While there have been many advances in patent searching techniques, there has been very little improvement in the art of broad searches such as ones that would be performed by academics, economists, or those looking for a general overview of a technological field. Atkinson (2008) discusses how little the methods for searching the patent database have changed in recent times, stating that

*'The Basic Way to search (mostly Boolean in logic structure, even if natural language has been used as a nexus) has changed little since ... 25 years ago)... We have far more databases available, more beautiful and comprehensive results display, but getting those hits still relies on set theory and exclusion.'*

An important case in the study of technological development is the work done by Trajtenberg (1987) in his analysis of computed tomography (CT) patents. Trajtenberg describes the use case of his set of 456 CT patents and importantly establishes that higher cited patents have higher value. To establish his patent set, he used his extensive case study of the CT industry (companies, installations in hospitals, inventors, etc.) to supplement a word search to find patents. He also read all the abstracts in his patent set to exclude inappropriate patents. Trajtenberg describes his method as one of trial and error in which he uses a variety of different aspects of the patent including the classification codes, assignees and regular Boolean search:

*'The search for patents in a particular product field or industry can be done in a variety of ways: using key words pertaining to the product in question that may appear in the title and/or in the abstract, identifying a small set of relevant patent classification codes, locating assignees (typically firms) that are known to operate in the field, etc. Needless to say, there isn't a well-defined method that would deliver with certainty* ***all*** *the patents in a given field, and* ***only*** *those.' (Emphasis his)* (Trajtenberg, 1987)

While Trajtenberg's method resulted in a patent set that is certainly more appropriate for his purposes than any others yet demonstrated, it is not clear that it can be reproduced in other technological domains and in fact the approach has not yet been applied anywhere else. It requires extensive knowledge of terminology in the field as well as information about relevant firms and he even detailed every installation of CT during his search, which was limited to 1971 to 1986. Our aim is to create a repeatable and simpler to use method in order that a user of the HKC method can quickly and easily compare patent sets across many technical fields of interest over a longer period of time than Trajtenberg considered. We do not anticipate any such simple method to reproduce what Trajtenberg retrieved, as much context he used is lost; however, a supplemental procedure should have value nonetheless.

There is a clear need for developing a supplemental approach as was pointed to by Atkinson (2008) when she discussed the "need for growth away from reliance on words and language and to draw in tools having defined quantitative values such as patent classes." The result of such an effort would then allow for an objective, repeatable methodology for selecting a set of patents that can then be analyzed to explore theories of technological change. The next sub-section will review several ways that patents have been analyzed for this purpose.

#### 2.3.3.2. Patent Analysis

Analyzing patents is essentially a 'Big Data' problem that asks how can meaningful information be extracted from exceedingly large amounts of data. This is exactly what many researchers have been attempting to do when they develop patent based metrics. In one of Manuel Trajtenberg's (1990) early papers, discusses some of the more basic patent measures, starting with simple patent count (SPC):

> *The body of evidence that has accumulated since Schmookler (1966) indicates fairly clearly that SPC are closely associated with the input side of the innovative process, primarily with contemporaneous R&D expenditures in the cross-sectional dimension (Griliches, 1984).* (Trajtenberg, 1990)

Trajtenberg adds evidence from his own work that the simple patent count in a particular technological field is a good proxy for the input variables of technological change (i.e. research effort or R&D). However, simple patent counts have proven to be less successful when attempting to measure the outputs of innovative effort:

> On the other hand, the few attempts to relate those counts to value indicators (e.g., the market value of innovating firms) have been largely unsuccessful. (See, for example, Griliches et al., 1988) ... those findings are hardly surprising, considering that patents vary enormously in their technological and economic significance. Thus, the mere counting of patents at any level of aggregation cannot possibly render good value indicators: simple patent counts assign a value of one to all patents by construction, whereas their true values exhibit a very large variance. (Trajtenberg, 1990)

While using simple patent counts as a measure for innovative outputs did not prove to be successful, Trajtenberg was able to discover that citation-weighted patent counts did correlate significantly with the value of an invention.

> Thus... citations are more informative of the value of innovations per se, rather than of the size of the market for the products embedding those innovations...The findings ... suggest that patent citations may be indicative of the value of innovations and, if so, that they may hold the key to unlock the wealth of information contained in patent data. (Trajtenberg, 1990)

This relatively simple idea that the number of citations that a patent receives is indicative of the value of the invention is another one of the underlying theories that form the basis of the research in this thesis. Trajtenberg mentions that part of the support for the patent citation relation to value can be found in literature from the Office of Technology Assessment and Forecast, which is a part of the United States Patent and trademark Office.

> During the examination process, the examiner searches the pertinent portion of the "classified" patent file. His purpose is to identify any prior disclosures of technology. . . which might anticipate the claimed invention and preclude the issuance of a patent; which might be similar to the claimed invention and limit the scope of patent protection ... ; or which, generally, reveal the state of the technology to which the invention is directed. If such documents are found they are made known to the inventor and are "cited" in any patent which matures from the application ... Thus, the number of times a patent document is cited may be a measure of its technological significance. (Office of Technology Assessment and Forecast, 1976, p. 167)

Trajtenberg continues to add practical theoretical support behind his argument:

> Moreover, there is a legal dimension to patent citations, since they represent a limitation on the scope of the property rights established by a patent's claims, which carry weight in court. Equally important, the process of arriving at the final list of references, which involves the applicant and his attorney as well as the examiner, apparently does generate the right incentives to have all relevant patents cited, and only those. (See Campbell and Nieves (1979).) The presumption that citation counts are potentially informative of something like the technological importance of patents is thus well grounded.

While the use of citations to define the importance of a particular invention is one of the most fundamental theories of this analysis, there are been many other metrics that extract information from the patent data. One particular example that will be used in this research is a metric that links patent data to the scientific basis of the invention as shown in Equation 6.

$$SCIENCE_i = \frac{NPCITES_i}{NPCITES_i + NCITED_i} \quad \text{(Equation 6)}$$

The metric finds SCIENCEi, which is a measure of the reliance on basic science for a patent, and is defined as the number of citations to non-patent literature (NPCITESi) to the total number of non-patent and patent citations (NCITEDi). The non-patent citations are almost always scientific journal articles, working papers or conference proceedings, thus represent basic scientific knowledge as opposed to the technical knowledge embodied in a patent. This measure is later supported by the fact that 'university patents do rely relatively more on non-patent (i.e. scientific) sources than corporate patents,' (Trajtenberg et al, 1997)

An example of the many different patent-based metrics that have been created in the past is found in the following list that includes the names of metrics for one study performed by Hall et al (2001).

***Original Variables:***

*1. Patent number*

*2. Grant year*

*3. Grant date*

*4. Application year (starting in 1967)*

*5. Country of first inventor*

*6. State of first inventor (if U. S.)*

*7. Assignee identifier, if the patent was assigned (starting in 1969)*

*8. Assignee type (i.e., individual, corporate, or government; foreign or domestic)*

*9. Main U.S. patent class*

*10. Number of claims (starting in 1975)*

***Constructed variables:***

*1. Technological category*

*2. Technological sub-category*

*3. Number of citations made*

*4. Number of citations received*

*5. Percent of citations made by this patent to patents granted since 1967*

*6. Measure of "generality"*

*7. Measure of "originality"*

*8. Mean forward citation lag*

*9. Mean backwards citations lag*

*10. Percentage of self-citations made –upper and lower bounds*

*(Hall et al, 2001)*

There are many other types of patent metrics that have been developed over the years, many of which are discussed in greater depth in Chapter 3 of this thesis.

#### 2.3.3.3. Limitations of Patents

While there are many reasons why patents are an attractive data source, they have limitations that should be considered when being used. In their large patent-based study, Trajtenberg et al (1997) provided two limitations of patents:

These data have, however, two important limitations: first, the range of patentable innovations constitutes just a sub-set of all research outcomes and second, patenting is a strategic decision and hence not all patentable innovations are actually patented. (Trajtenberg et al, 1997)

They mention that not all inventions can be patented by the definition of patents:

Maxwell's equations could not be patented since they do not constitute a device (ideas cannot be patented); on the other hand, a marginally better mousetrap is not patentable either, because the innovation has to be non-trivial. Thus, our measures would not capture purely scientific advances devoid of immediate applicability, as well as run-of-the-mill technological improvements that are too trite to pass for discrete, codifiable innovations. (Trajtenberg et al, 1997)

And that sometimes, it does not make economic sense to patent all inventions, even if they fall under the patentability criteria:

The second limitation is rooted in the fact that it may be optimal for inventors not to apply for patents even though their innovations would satisfy the criteria for patentability. For example, until 1980 universities could not collect royalties for the use of patents derived from federally funded research. This limitation greatly reduced the incentive to patent results from such research, which constitutes about 90% of all university research. Firms, on the other hand, may elect not to patent and rely instead on secrecy to protect their property rights (there is a large variance across industries in the reliance on patents versus secrecy: see Levin et al, 1987). Thus, patentability requirements and incentives to refrain from patenting limit the scope of measures built on patent data. (Trajtenberg et al, 1997)

Beyond some of the basic limits of patents due to patentability criteria, there are other factors that should be considered, such as ones having to do with the temporal nature of patents. For some metrics, truncation can be an issue, as more recent patents will not have had enough time to be patented, and can possibly be labeled incorrectly as unimportant. This problem was documented by Hall et al (2001):

> On the other hand one has to be mindful in that case of the truncation problem: as the time series move closer to the last date in the data set, patent data timed according to the application/publication date will increasingly suffer from missing observations consisting of patents filed in recent years that have not yet been granted. (Hall et al, 2001)

Another issue that can cause issues in some of the metrics is the ever-increasing number of patents. This issue could impact metrics that depend on the number of patents at a certain time and thus can be corrected by normalizing for different time periods (usually decades or years).

Ultimately, Trajtenberg et al decided that the limitations we outweighed by the advantages in the use of patents for technological analysis.

It is widely believed that these limitations are not too severe, but that remains an open empirical issue. (Trajtenberg et al, 1997)

The authors of this research take the same position and believe that the overall benefits of using patents as proxies for inventions (generalizability, objectivity, quantitative and qualitative nature) outweigh the limitations of patents as a data source.

# Chapter 3: Methodology

The main focus of this research can be summarized into 4 overarching goals.

1. To develop reliable technological improvement curves and the related exponential improvement rates of many different technological domains

2. Select sets of patents that represent those technological domains for which there are reliable technological improvement rates that can then be analyzed to better understand the characteristics of the domains

3. Develop a methodology to unify the quantitative technological improvement trends and the patent based characteristics of the domains

4. Analyze the results of the comparison between the quantitative technological improvement trend and the patent based characteristics to improve our understanding of technological change and to increase our technological prediction capabilities

In this section, through the process of explaining the methodology for accomplishing goal #'s 1, 2 and 4, the main components of goal 3 will be introduced and detailed.

The methodology section will begin with a definition of a technological domain and how the 28 specific domains of interest were selected. This will be followed by the introduction of functional performance metrics (FPMs) and will include a review of the appropriate independent variables (horizontal axis) for measuring technological improvement. An in-depth examination of manufacturing progress will follow in which new and complex FPMs are defined for manufacturing technologies that have not been studied this way before.

The 2nd half of the methodology section will first describe the classification overlap method (COM) that is used to select a set of patents in a repeatable and objective manner. The COM will then be used to select patent sets for each of the 28 domains of interest. The next section will detail all of the domain specific markers (DPMs) that were extracted from the patent data. While lengthy, this section provides a novel set of methods that allows unification of quantitative technological trends and patents- arguably the two most important objective data types for technological change. A flow chart of the entire methodology is shown in figure 17. Finally, a short case study comparing the TIRs and DPMs of renewable energy domains will provide context for the large 28 domain experiment.

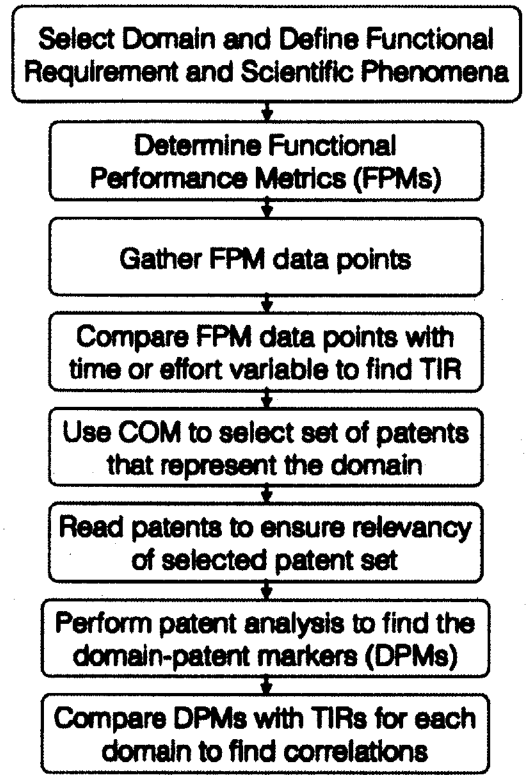



**Figure 17: A flow chart of the sequential steps in the methodology for the cross-domain comparison of Technological Improvement Rates and Domain Patent Markers**

## 3.1. Domains

One of the least repeatable and generalizable aspects of technological change research is the selection of the unit of analysis. Many possible levels of the units of analysis are possible for understanding how technologies change over time and are shown on a continuum in Figure 18. There are some studies that have examined specific inventions (or unspecified sets of inventions) at specific times, as was demonstrated by Tushman and Anderson's (1986) list of technological discontinuities or Girifalco's (1991) list of innovations since the 18th century that were discussed in Chapter 2.2.7. Others, such as Solow (1956), have studied all of technology in general, in an attempt to explain economic growth that is not caused by additional labor or capital. More commonly, researchers attempt to study specific technologies. Studying technologies at this intermediary level eliminates the subjectivity and lack of breadth of selecting individual inventions, while allowing for more specificity and deeper analysis than viewing technology as a whole.

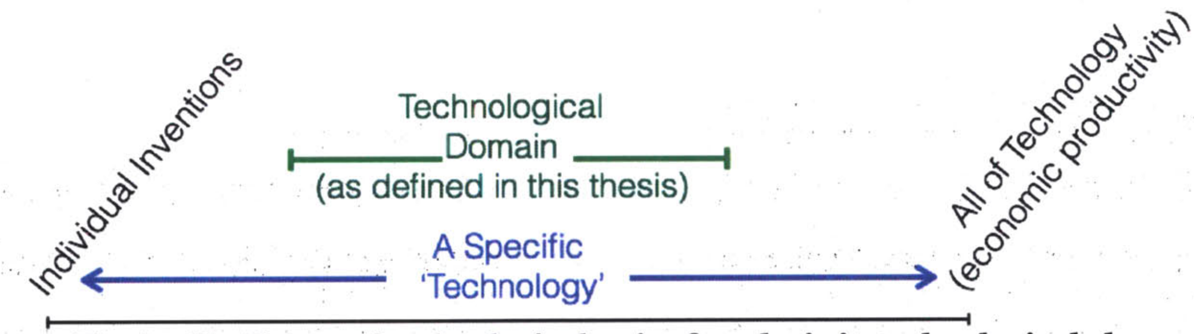



**Figure 18: Range of technological unit of analysis in technological change research and a technological domain as defined in this thesis.**

Although the level of analysis has been narrowed down slightly, there is still much flexibility in the term "a technology". While there are many attempts at answering this question, Dosi (1982) and Arthur (2007) address this problem quite well, and are the basis for the unit of analysis in this thesis. Arthur posited that any technology has two main components. The first component is that any technology is 'a means to fulfill a human purpose.' Examples of purposes are 'to power an aircraft', or 'to sequence a DNA sample,' or to 'generate electricity.' The second component of technology is that it must take advantage of an effect or phenomenon. This effect could be something like the conversion of light to electrons through the photoelectric effect, or the mathematical principles that govern radio waves; the effects do not necessarily need to be physical, they can be scientifically, mathematically, or even socially based. Thus Arthur's definition of a technology is:

*'a technology is a means to fulfill a purpose, and it does this by exploiting some effect.' (Arthur, 2006)*

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

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

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

In determining the unit of analysis for this thesis, many of the underlying concepts behind Dosi and Arthur's definitions are maintained, while the definition of technology is slightly modified to fit the purpose of this research. First, due to the significant and different uses of the term 'technology', the term that will be used in this thesis is ***Technological Domain (TD)***, which provides clear differentiation from the other uses of the term 'technology'.

A *technological domain* can be defined as: **The set of artifacts that fulfill a specific generic function utilizing a particular, recognizable body of knowledge.**

This definition is more specific in terms of the set of artifacts (which includes systems, processes and algorithms as well as devices) than Arthur's use of the term 'means.' Additionally, the term purpose is less ambiguous when it is described as a specific generic function. The specific generic function will be a main area of focus in later sections of this paper that define specific performance metrics for a particular technological domain. The precision in this term provides more clarity about the relationship between a domain and their performance characteristics and links the technological domain to its economic purpose. Finally, the term 'some effect' has been replaced by 'a particular, recognizable body of knowledge,' in an attempt to more closely link the technological domain with the underlying knowledge that it is based upon and reduced uncertainty about unknown effects that are not yet considered 'knowledge' that may crosscut several technological domains.

It is also important to note the areas that the definition is intentionally non-specific. The two terms to take notice of are the '**set** of artifacts..' and '...a recognizable **body** of knowledge.' These two terms allow for a technological domain to be as broad as 'semiconductors' or as narrow as 'industrial stereolithography 3D printers'. The fact that this definition does not require a certain level of specificity makes it more flexible and able to represent a large set of potential technologies. Another benefit of this flexibility is that it is likely impossible to create a specific set of technological domains that map the entire space of technology, and there are nearly an infinite number of possible ways to construct a total technology map. This flexible definition of a technological domain allows for the scale and scope of a domain to be adapted to the goals of the specified research. The range of the technological domain as defined in this thesis is shown in schematically figure 18.

### 3.1.1. 28 Technological Domains

In total, 28 technological domains (TDs) were studied in this thesis. The TDs were selected for analysis based upon a variety of factors. The first and most limiting factor is data availability, without data on performance or recent patents, it is impossible to perform the analysis required for the cross-domain comparison of technological improvement rates and patent derived characteristics. The next factor was looking for TDs that represented a broad range of both generic functions and bodies of knowledge in order to increase the potential generalizability of the study and allow for further cross-domain testing. For example, one hypothesis about technological improvement posits that the differences in improvement rates could be driven by the type of technology (electrical vs chemical vs mechanical) (Koh and Magee, 2008); studying a wide variety of domains will allow such ideas to be explored. Another factor that contributed to the selection of these specific 28 TDs was the general scope of the fields. As was mentioned in the previous subsection, the definition of a TD allows for a wide range of scopes, and therefore including domains across a portion of the spectrum shown in Figure 18 is important for generalizing the results of the study to many different technologies. A more quantitative measure of scope will be introduced when the patents sets for each TD are defined.

Table 4 shows the 28 domains that were selected for analysis in this thesis within their functional performance category as is derived from Koh and Magee (2006, 2008). The first row of the table is the operand on which the domain acts, and the first column of the table shows the operation that the technological domain performs.

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

**Table 4: The 28 domains studied in the thesis classified by functional technological classifications with operands and operations, adapted from Koh and Magee (2006, 2008)**

The previous work done on this functional technology classification system shows that the 9 types of classifications represent a relatively complete overview of all possible technologies. The 28 domains analyzed in this thesis fall into 8 out the 9 (with matter storage being the exception) possible operand-operation classifications and thus represent a very wide range of technological functions.

A few of the domains were selected to allow for comparison to one another, such as solar photovoltaic and wind turbines. Additionally the manufacturing domains were selected specifically due to the lack of prior research on quantitative technological trends in manufacturing domains and their addition is an important contribution of this thesis. Details regarding the selection of the manufacturing TDs will be covered in depth in section 3.3.

## 3.2. Determining Technological Improvement Rates and Their Reliability

The next step in the methodology for the large cross-domain comparison of technological improve and patent characteristics is to determine the technological improvement rate (TIR) that represents the performance improvement for a specific generic function that the technological domain is accomplishing. This is done by first constructing a functional performance metric (FPM) that is a measure of the generic function for a TD and includes the factors that affect the purchasing decision for the technology. Next, data points that measure the FPM are collected over a range of time and a technological improvement rate is determined by an exponential regression vs time, as was introduced by Moore and discussed in section 2.1 of this thesis. Finally, the TIR is statistically analyzed to examine robustness and reliability. The TIRs contain a significant amount of information about the past performance and future potential of a technological domain. One of the main goals of this research is to understand why differences in TIRs between domains exist, thus the creation and validation of the TIRs is a critical step in this research.

### 3.2.1. Constructing the Functional Performance Metric (FPM)

The functional performance metric is the measure used to assess the performance of the TD. For example, an FPM for electrochemical batteries could be kWhr/kg - or the energy density. This measure is tied to the purchasing decision of the specific TD; when the FPM improves, a purchaser would pay less (in this case have a lighter energy source) for or gain more value from the technology.

In developing FPM, it is important to consider all of the possible aspects of a technology that the consumer may value. This includes the general value that they will receive from the product as well as any costs to the purchaser (which encapsulate more than the purchase price). For the case of electrochemical batteries, a consumer would be looking to access stored kWhr, which would be the value of the technology, and a cost to the consumer would be mass (kg), volume (L), and price ($). One way to think of an FPM is a trade-off surface that is constrained by cost and performance and technological change and that can be defined as:

The technological change of a product with many elements of performance, whose characteristics and prices change over time, is the change in total factor inputs required to produce the product, holding its characteristics (performance or output) constant. (Alexander and Mitchell, 1985).

As shown in Figure 19 This definition then states that a technology can improve when either (A) the cost of a technology decreases without any decrease in performance or (B) the performance of the technology can improve without an increase in cost. There is also the option for (C) where the performance increases and the cost decreases at the same time.

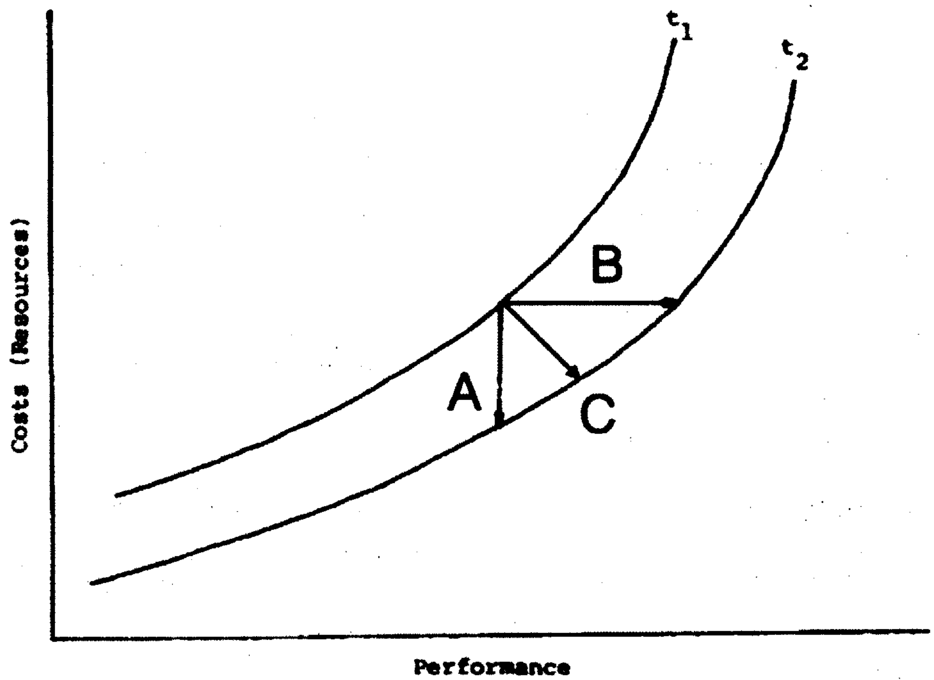



**Figure 19: Two Possible ways for a technology to improve (A) by decreasing cost at a constant performance and (B) increasing performance at a constant cost - adapted from Alexander and Mitchell (1985)**

It must be reiterated that the term 'cost' in Alexander and Mitchell's definition is not one only of monetary purchase price ($), but also represents the costs mentioned above such as additional weight or volume. Other domains can see costs such as noise, energy consumption, surface area, etc.

In many cases, the purchase price, in USD, is an important variable, however it can introduce a considerable amount of short-term (and possibly long-term due to depletion effects) noise to the data due to exogenous effects that are not related to the technological development such as government subsidies and changing trade patterns or depletion of natural resources.

In order to evaluate both value-creating parameters and total costs (weight, volume, price, etc.) together, it is important to construct the functional performance metric properly to take into account their opposing effects on consumer utility. Because of this, FPMs are constructed so that their value increases with an increase in customer utility. When constructing an FPM, the parameters that are positively correlated with consumer value are placed in the numerator of the FPM and the negatively correlated parameters (such as price) in the denominator of the FPM. Occasionally, a value-creating parameter is negatively correlated with consumer value and should be placed in the numerator of the FPM. For example, when considering a manufacturing technology, often times a lower precision value is actually better, therefore the FPM should be constructed using 1/precision. In fact, the proper construction of an FPM must include proper 'trade-offs' in order for the performance of a TD to be measured properly. Figure 20 shows a plot for information storage with (a) the logarithm of a simple FPM with time as a tradeoff; millions of instruction per second (MIPS) vs time and (b) a more complete FPM, the logarithm of MIPS/$ vs time.

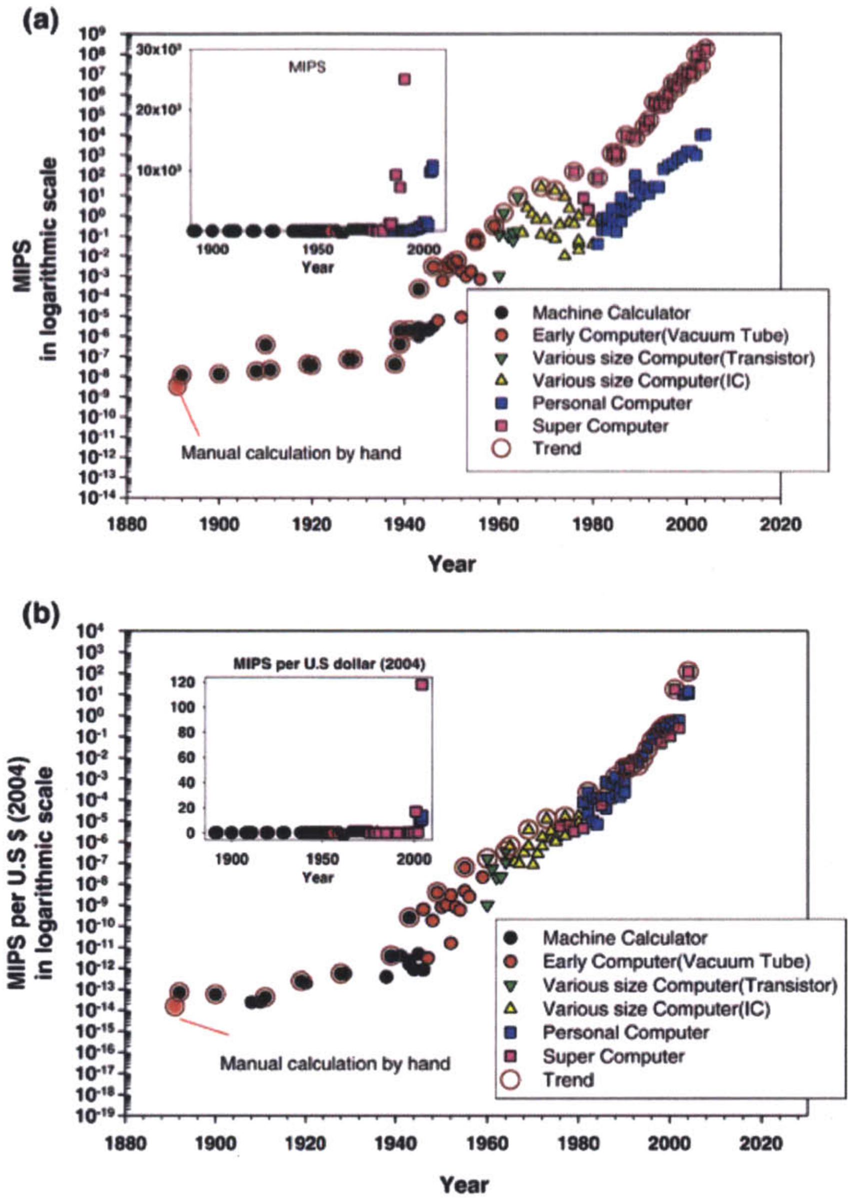



**Figure 20: Increase of computational power over time, as measured by (a) logarithm of MIPS and (b) logarithm of MIPS/$ – adapted from Koh and Magee (2008)**

The inclusion of price in the functional performance metric reduces the scatter in the trend. The simple FPM (a) shows only the maximum MIPS for a particular device, with no cost taken into account, and it is logical to assume that a very expensive super computer would be able to perform many more instructions per second than a relatively inexpensive personal computer. The inclusion of price brings the points closer together and portrays a more accurate picture of the improvement of computation devices than the simpler FPM.

Any variables that may affect consumer value that are not included in the FPM for a TD are considered 'omitted variables' as mentioned by Alexander and Mitchell (1985). In their paper, Alexander and Mitchell claim that omitted variables are too significant of a problem to allow for proper measurement of technological change. There has been much research since that paper to show that while omitting a significant number of variables detracts from the reliability of the technological change measurement, using FPMs to measure technological is certainly possible. It is important to keep in mind, however, all of the aspects of a particular TD that may be important to a purchaser.

In the ideal case there would be no omitted variables and the measure would be considered a 'complete' FPM, however this is practically impossible as there are always unknown consumer preferences for products, often times ones that aren't even being considered by the designers. An example of an omitted variable for the electrochemical battery FPM that was listed above is the maximum power output of a particular battery. While it may seem logical then, to continue to add even the most esoteric factor that may influence the purchasing decision for a TD, this must be balanced with the need for data points in which the FPM has been measured. Often times the most complete FPMs will have very few data points where all of the individual metrics are measured together.

There are many ways to construct FPMs for each TD, therefore in many cases each TD has a number of FPMs that can represent the improvement of the technology. This is similar to what Christensen (1992A) discussed when mentioning the performance metric for a technology in one market vs another. It is important to consider many different combinations of the parameters that make up each FPM. For example, the improvement of electrochemical batteries can be measured by the following FPMS: kWhr/kg, kWhr/L, kWhr/$, kWhr/(kg*$), kWhr/(L*$). It is important to note that not all of these FPMs will have sufficient data to construct a technological improvement curve (TIC).

In summary, the practice of constructing a set of FPM(s) for a TD is an exercise in dimensional analysis with the goal of creating the most complete metric for which their exists enough data to compare across time. After compiling the FPMs, each one should be combined with their respective TD to create a set of domain metric pairs (DMPs) for each TD. The complete list of DMPs for each TD is described later in section 4.1 of this thesis.

### 3.2.2. Constructing the Technology Improvement Curve

Finding the functional performance metric data to create a technology improvement curve (TIC) is one of the most difficult, tedious, and time-consuming processes of this thesis and is central to understanding how technology changes over time. In order to get a reliable understanding of how a technology has changed, it is important to get as many data points as possible and to cover as wide of a time-frame as possible.

Additionally, as was mentioned in the previous section, some of the FPMs include a number of different parameters for each TD. In order to construct a reliable FPM, it is important to find data points that include as many different parameters as feasible. For example, in order to evaluate manufacturing technologies, multiple parameters must be found for each point and include: precision, cost, speed, flexibility, weight, and volume. Finding 5 or 6 different parameters across a wide range of time can be a significant task and requires searching a wide variety of information sources as will be discussed in the next sub-section.

Although finding the FPM data is a difficult task, there are a number of different data sources to consider in doing so. These data sources will be covered in the following subsections of the thesis.

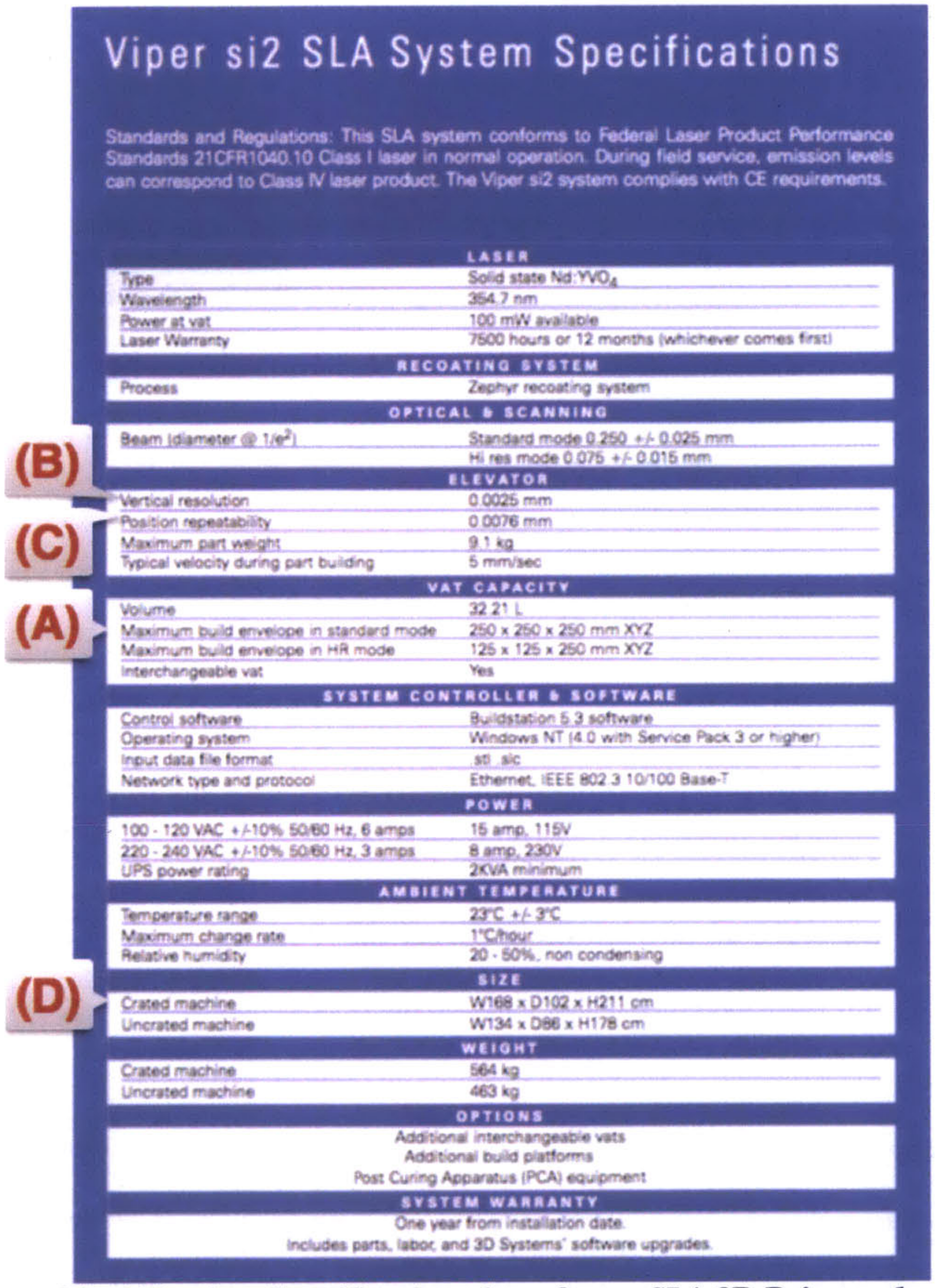
#### 3.2.2.1. Data from the Producers of the Technology

Very few producers and companies spend a significant amount of time and resources in keeping track of the historical performance of their products, which makes finding the FPM data difficult. Finding current data for the metrics is easier because many companies do track the current performance of their products in an effort to benchmark those against the competition. When speaking with representatives from a number of companies, they explained that there is little value and significant cost to keeping a well-organized archive of product performance. While this research is not consistent with the idea that historical performance is of little value, there is no questioning the fact that maintaining data in an archive or museum can be expensive and time consuming, and thus is usually only done sporadically by the largest companies.

Even in the case of some large companies who have kept this information, the data is often poorly organized and thus requires the reading of a significant amount of information to extract the few data points of interest. Another problem with using data from companies is that in some cases the data is confidential and is used as a competitive advantage. This was the case in the wind turbine industry for a number of years, and is still the case with some of the more specific details regarding the production of carbon nanotubes. Due to this, many of the data points that come from companies that produce the technologies come from anonymous interviews in which the specific data points that are given are often 'rough estimates' and are required to be reported anonymously so as to maintain competitive advantage. Nonetheless, information on technological performance that comes directly from companies, even if it comes with wide error ranges, can be very useful when combined with data points from other sources because of the immediate linkage to the producer of the technology.

#### 3.2.2.2. Product Specifications

While collecting information from companies directly can be a difficult task that nets a wide error range for the data points, often companies release specification sheets for their products. These 'spec sheets' provide a significant amount of information on each product and often contain many of the parameters that make up the FPMs, making them a good source of information for constructing TIRs. Figure 21 shows an example spec sheet of a stereolithography 3D Printer. In it there are many parameters regarding the performance and technical characteristics of 3 different 3D Printers. These specifications show the net build volume, the resolution, accuracy, machine size and weight which can be used to construct an FPM for these machines that includes both valued traits (build volume, layer thickness) and costs (machine volume and weight).



## Viper si2 SLA System Specifications

Standards and Regulations: This SLA system conforms to Federal Laser Product Performance Standards 21CFR1040.10 Class I laser in normal operation. During field service, emission levels can correspond to Class IV laser product. The Viper si2 system complies with CE requirements.

| | |
|---|---|
| **LASER** | |
| Type | Solid state Nd:$YVO_4$ |
| Wavelength | 354.7 nm |
| Power at vat | 100 mW available |
| Laser Warranty | 7500 hours or 12 months (whichever comes first) |
| **RECOATING SYSTEM** | |
| Process | Zephyr recoating system |
| **OPTICAL & SCANNING** | |
| Beam (diameter @ $1/e^2$) | Standard mode 0.250 +/- 0.025 mm<br>Hi res mode 0.075 +/- 0.015 mm |
| **ELEVATOR** | |
| (B) Vertical resolution | 0.0025 mm |
| (C) Position repeatability | 0.0076 mm |
| Maximum part weight | 9.1 kg |
| Typical velocity during part building | 5 mm/sec |
| **VAT CAPACITY** | |
| Volume | 32.21 L |
| (A) Maximum build envelope in standard mode | 250 x 250 x 250 mm XYZ |
| Maximum build envelope in HR mode | 125 x 125 x 250 mm XYZ |
| Interchangeable vat | Yes |
| **SYSTEM CONTROLLER & SOFTWARE** | |
| Control software | Buildstation 5.3 software |
| Operating system | Windows NT (4.0 with Service Pack 3 or higher) |
| Input data file format | .stl .slc |
| Network type and protocol | Ethernet, IEEE 802.3 10/100 Base-T |
| **POWER** | |
| 100 - 120 VAC +/-10% 50/60 Hz, 6 amps | 15 amp, 115V |
| 220 - 240 VAC +/-10% 50/60 Hz, 3 amps | 8 amp, 230V |
| UPS power rating | 2KVA minimum |
| **AMBIENT TEMPERATURE** | |
| Temperature range | 23°C +/- 3°C |
| Maximum change rate | 1°C/hour |
| Relative humidity | 20 - 50%, non condensing |
| **SIZE** | |
| (D) Crated machine | W168 x D102 x H211 cm |
| Uncrated machine | W134 x D86 x H178 cm |
| **WEIGHT** | |
| Crated machine | 564 kg |
| Uncrated machine | 463 kg |
| **OPTIONS** | |
| Additional interchangeable vats<br>Additional build platforms<br>Post Curing Apparatus (PCA) equipment | |
| **SYSTEM WARRANTY** | |
| One year from installation date.<br>Includes parts, labor, and 3D Systems' software upgrades. | |

**Figure 21: A Product Specification sheet for an SLA 3D Printer showing the (A) build volume, (B) layer thickness, (C) accuracy, and (D) system dimensions, all of which can be used together to form an FPM for measuring the performance of SLA 3D Printers – adapted from (www.3dsystems.com, 2014).**

Spec sheets such as these are released often alongside the announcement of a new product or technology, thus many spec sheets exist over time. While some of the older spec sheets can be difficult to find, in the age of the Internet and digital information, spec sheets can be very useful in constructing technological improvement rates.

While spec sheets span long periods of time and contain a large amount of technical and performance data concerning a technology, their purpose is not to track technological improvement rates, rather they are designed as an advertising component to sell more of that product. Given this, it is important to take the data included in the spec sheet with a grain of salt, as sometimes the specifications are difficult to test and thus are slightly exaggerated. This case can be often be identified in industries where a set standard of performance seems to be nearly identical for a large number of products. An example of this is in the consumer 3D printing market where almost all of the machines claim to have very similar layer thickness values of 0.1mm (www.3ders.org, 2014) as shown in Figure 22.

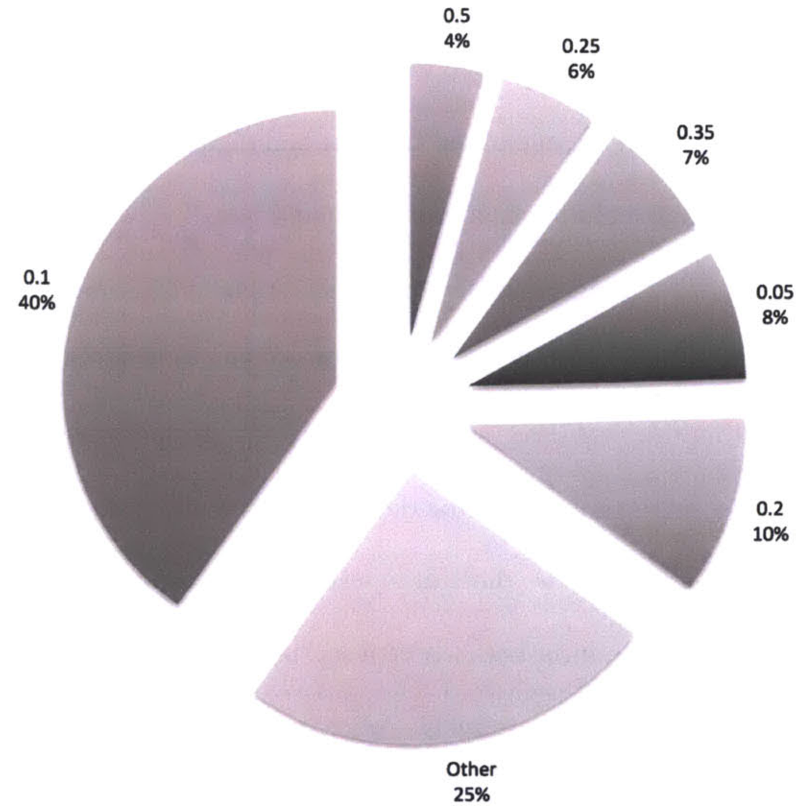



**Figure 22: Distribution of layer thickness specifications for consumer 3D printers**

While the actual values are likely to be close to the 0.1mm, it is likely that their is more variation than is stated by the spec sheets and that an unwritten 'standard' has been accepted for the minimum layer thickness that a consumer will accept. Despite these flaws, spec sheets remain an acceptable source of information for construction technological improvement rates.

### 3.2.2.3. Trade Magazines

One great source of finding technological specifications over time is trade magazines, where the producers of the technologies often advertise new products and consumers of the technology research new products. Trade magazines have similar advantages to spec sheets in that they are published on a continual basis and thus long time frames of information exist. Additionally, trade journals often include the most relevant parameters to measure the performance of a technology, and thus are good sources to find data for constructing TIRs. Beyond the advertisements in the trade journals, the journals often have experts (writers and editors) who monitor the field closely and write columns on the changes in the industry and new artifacts or processes that may be introduced. The combination of the articles and the advertisements provides a source of qualitative and quantitative information to help construct TIRs.

Potential downsides of trade magazines are very similar to that of spec sheets. Trade magazines are often funded by selling advertisements and thus the information is designed to sell products and should be recognized as so when constructing the TIR.

### 3.2.2.4. Scientific Literature

While data gathered from industrial sources has the advantage of close proximity to the creator, it is not published in a peer-reviewed format. Scientific journal articles and conference proceedings are a peer-reviewed source of information that in some cases contains FPM data. The main benefit of scientific literature is the assumed accuracy. Additionally, in some cases one source can contain multiple FPM data points, which can populate a significant portion of the data required for a TIR. Scientific literature also often includes a wider range of data that can be attributed to more than one producer, reducing the possibility that the particular producer is an outlier in performance for a particular TD.

While the peer-reviewed aspect contributes to the validity of the data, it is often necessary to look closer at the source(s) from which the data is taken from. In many cases, the scientific articles have simple gathered data from the sources listed above. In some cases, the author has ran specific experiments to complement (or supplant) the industry given data, which is useful in that it provides another independent source. Finally, many journal articles and conference proceedings can be difficult to access for non-academics. Overall, peer-reviewers sources are some of the most desirable to use when constructing a TIR.

#### 3.2.2.5. Industry Reports

Industry reports are documents compiled on the state of an industry by either private firms or sometimes government agencies. Like scientific literature, industry reports often include multiple data sources in one location, significantly reducing the search time for populating a TIR. Industry reports also generally contain aggregate data for an industry and thus give a better *average* picture of the industry rather than the company-specific information contained in some of the other sources. These reports are often compiled by skilled and well-compensated people who work for companies with close ties to the industries that they are covering - increasing their apparent reliability as they should eliminate unreasonable inconsistencies. Often times a significant portion of the data used to contract a TIR may be from a set of industry reports.

While generally a reliable and well-populated source of information, the private-firm industry reports can be difficult to procure and often present their information with few references. The reports that are compiled by private firms are often very expensive and difficult to access (sometimes even for academics). Additionally, due to the proprietary nature of their information and sources, the private-firm industry reports often do not contain any references to their data. Thus, their reliability is not considered as high as those by public agencies.

Industry reports filed by public agencies (such as the U.S. department of energy) are often free and easily accessibly on the internet, however they, more-so than the other sources, contain data that is a large aggregation of the industry, and thus should be considered as so. In particular, the average data is useful for constructing TIRs of complex FPMs, but can be less reliable when attempting to chart the best performing example in a TD for a given year.

In summary, industry reports are generally a good place to start the search for FPM data points, but often a specific DMP has not been tracked in an industry report and thus the other data sources must populate the TIR.

### 3.2.3. Calculating a TIR

After the performance data is collected, it must be compared with another independent variable. As was discussed in Section 2.1 of the thesis, while there are a number of different ways to evaluate technological improvement trends, the TIRs in this thesis are all based upon comparing performance with time.

Most FPM data is connected to a corresponding year such as a product release date, correspondence with a company, article publication date, etc. Thus, it is rather straightforward to compare the FPM data with time. As was mentioned previously, the FPMs should be constructed so that they increase as performance of the product improves or the cost decreases. The most common way of plotting the FPM with time is with the FPM on the vertical axis (dependent variable) and time on the horizontal (independent variable). When this is the case, the technological performance trend is expected to show an increase over time as is shown in Figure 23.

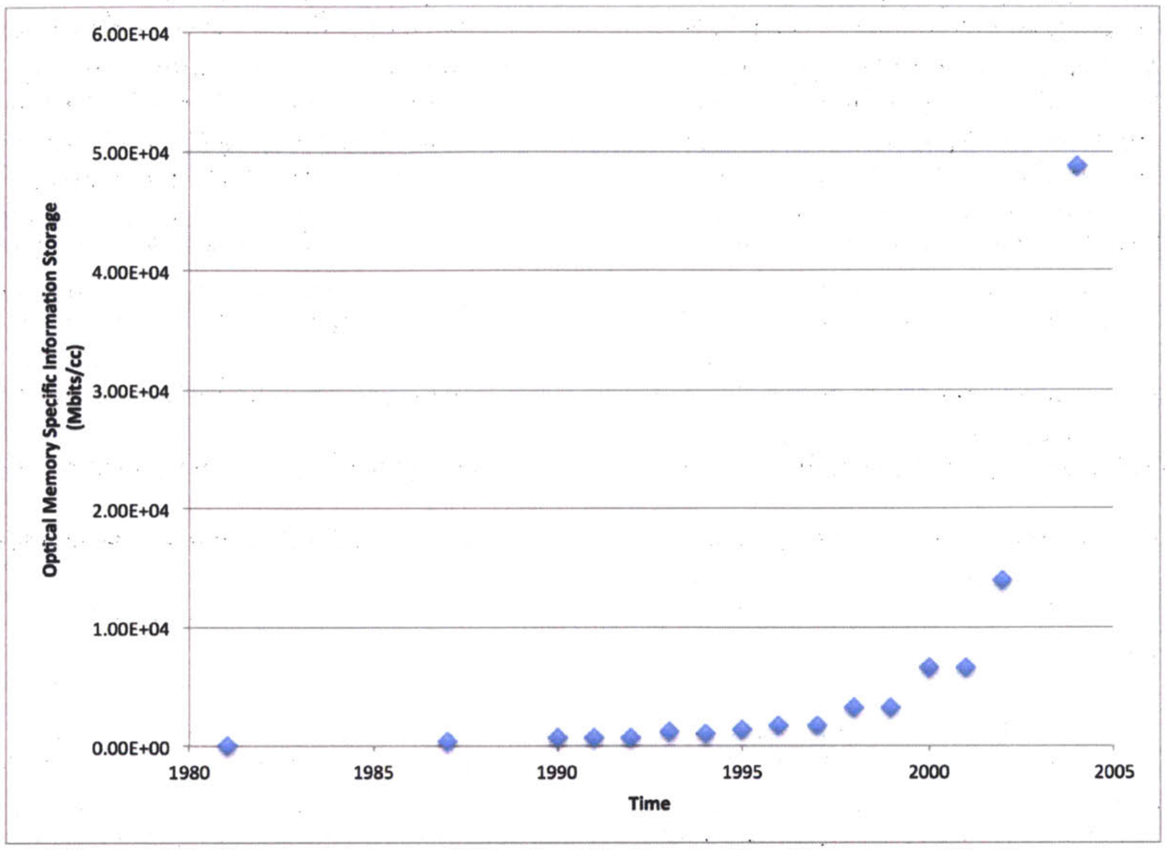



**Figure 23: Improvement of Optical Information Storage with time on a linear-linear scale, adapted from (Koh and Magee, 2006).**

When plotted in this manner, the FPM performance appears to improve exponentially with time, which is consistent with the prior literature mentioned in section 2. Due to this relationship, it is easier to observe the exponential improvement when the data is plotted on a log-linear graph as shown in Figure 24.

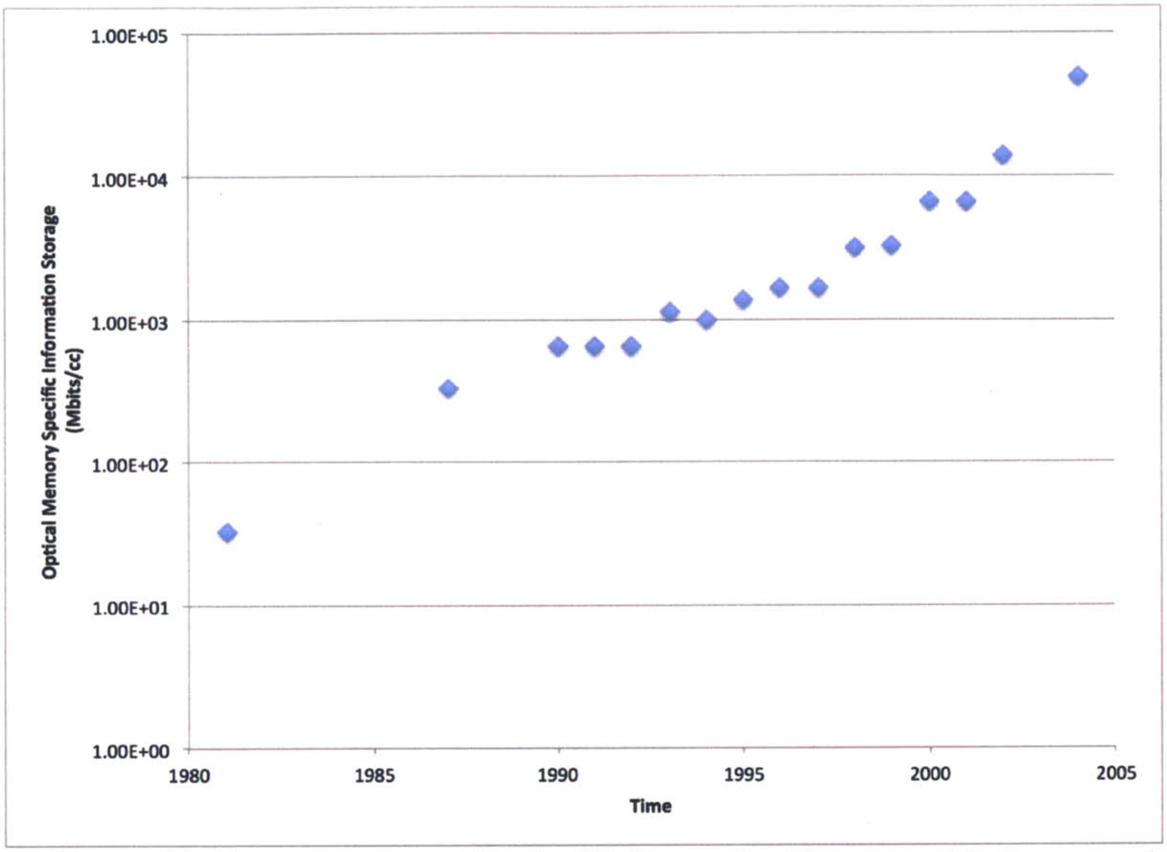



**Figure 24: Improvement of Optical Information Storage with time on a log-linear scale, adapted from (Koh and Magee, 2006).**

The trend now appears much more linear (but not perfect) when plotted on a log-linear graph, and thus indicates that the relationship between the FPM and time is fundamentally exponential. How well this description works can of course be mathematically determined and is shown by the exponential trend line in Figure 25.

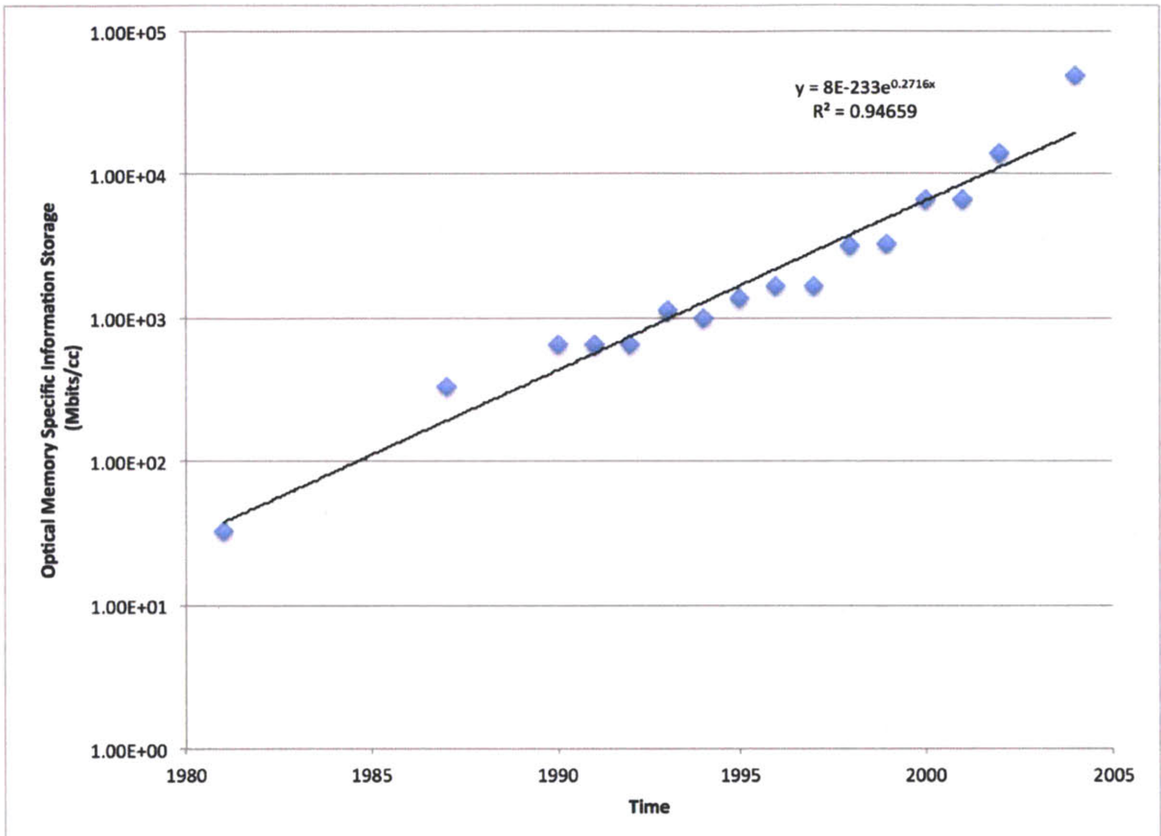



**Figure 25: Improvement of Optical Information Storage with time on a log-linear scale with an exponential trend line, adapted from (Koh and Magee, 2006).**

There are a few things to note about the trend line. First, the *trend line* (not the data points) appears perfectly linear on the log-linear graph, which shows the example of a perfectly exponential improvement; the closer the FPM data points are to being perfectly linear on the log-linear graph, the closer they are to being perfectly exponential. The next aspect to note is the form of the exponential equation, which should resemble equation 7 where $y$ is the FPM value, $c$ is a scaling factor, $x$ is the date (in years) and $k$ is the exponential improvement rate. It is important to note that this equation takes the same form (yet different variable names) as equation 2 that was discussed in section 2.1 of this thesis and is based off of Moore's Law.

$$y = ce^{kx} \quad \text{(Equation 7)}$$

<u>The exponential improvement rate $k$ is the technological improvement rate that is used for the main analysis portion of the research in this thesis.</u>

This value represents the constant percentage improvement of the FPM. An easy way to think about TIR is to compare them to bank account interest rates. A bank account that offers a higher interest rate will give you more return on your money than one that offers a lower rate. For example, if bank account 'A' offered a 10% interest rate, without any additional deposits, the money will double in a little over 7 years; if bank account 'B' offered a 30% interest rate, the money will double in approximately 2 years and 4 months. A quick way to calculate an *estimated* doubling time is the rule of 72, which states that in order to find the amount of time it would take to double an initial balance, simply divide 72 by the interest rate as shown in Equation 8.

$$t_{double} = \frac{72}{k} \quad \text{(Equation 8)}$$

It is also important to take note of the $R^2$ value of the fit of the data to the trend line. This value shows how close the data is to being perfectly exponential. A more specific definition of $R^2$ (AKA the coefficient of determination) is the percentage of total variation in Y can be described by the trend line.[1] The value of $R^2$ can vary between 0 (no fit between the data and the trend line) and 1 (the data is perfectly consistent with the trend line and therefore perfectly exponential). In practice the values are never as low as 0 or as high as 1, and for this research we strive for $R^2$ values that are higher than 0.8 and do not accept those less than 0.6. The next sub-section will explore more rigorous statistical testing of the TIRs.

### 3.2.4. TIR Reliability Measures

While most technological domains exhibit an exponential relationship with time, there is a wide variation between the TD and even some variation between the domain-metric pairs within a TD. While there does exist a noticeable variation in technological improvement rates between domain-metric pairs within a TD (intra-domain) the variation between different TDs (inter-domain) is much larger, as is shown in Appendix A of this thesis and thus most of the statistical tools for comparing the TIRs will focus on comparison with other TDs.

In the last subsection, the $R^2$ value was used as a simple method of determining the reliability of each TIR. This section will introduce several statistical methods for measuring the reliability of the TIRs as well as a method for selection of the most-complete reliable TIR, and the most-reliable complete TIR to be used for the large cross-domain study that is one of the main goals and innovations of this thesis. The statistical measures used to evaluate the TIRs are discussed in the following subsections.

#### 3.2.4.1. R-Squared

As was mentioned in the previous subsection, the $R^2$ value is a measure the percentage of deviation in the data points that can be explained by the exponential regression and can be calculated using equation 9.

$$R^2 = 1 - \frac{\sum_i (y_i - f_i)^2}{\sum_i (y_i - \bar{y})^2} \quad \text{(Equation 9)}$$

The TIR is the exponential coefficient as shown in equation 7 (*k*) therefore the $R^2$ is also a measure of how well the TIR actually reflects the data that was collected. This measure is also one that can be found in many of the other prior papers in the field of technological research and thus is one of the most important reliability measures used to evaluate the TIRs.

### 3.2.4.2. Number of Data Points

As was described in the data collection sub-section earlier, more data points is generally an indicator of a more reliable TIR. While this measures is also reflected somewhat in the $R^2$ value, it is useful to list it separately because the $R^2$ value can sometimes be misleading when there are few data points. For instance when there are only 2 data points, the $R^2$ value is 1, because any line can be perfectly fit between only 2 points.

Along with the number of data points, it is also important to consider the sources from which the data points were derived. More sources of data used to construct the TIR is indicative of a higher reliability due to the fact that multiple data sources usually indicates multiple independent observations.

Finally, the types of data sources that are consulted are also important to consider when measuring the reliability of the TIRs. As was mentioned in a previous sub-section, all data sources are not created equal and thus sources such as peer-reviewed scientific journal articles are likely to be more reliable than anonymous estimates from an interview with someone working directly for a producer of a technology.

The combination of the number of data points, the number of data sources and the type of the data sources provide critical quantitative and qualitative information about the reliability of the TIRs.

### 3.2.4.3. Range of Time Covered

While having many data points from many different sources is important, if they are all condensed in one small timeframe, it is less likely that the TIR is generalizable to longer timeframes. This is especially true in some of the cases with higher noise, because although long term exponential improvement is rather consistent, this is not true when considering very short time frames as is shown in Figure 26.

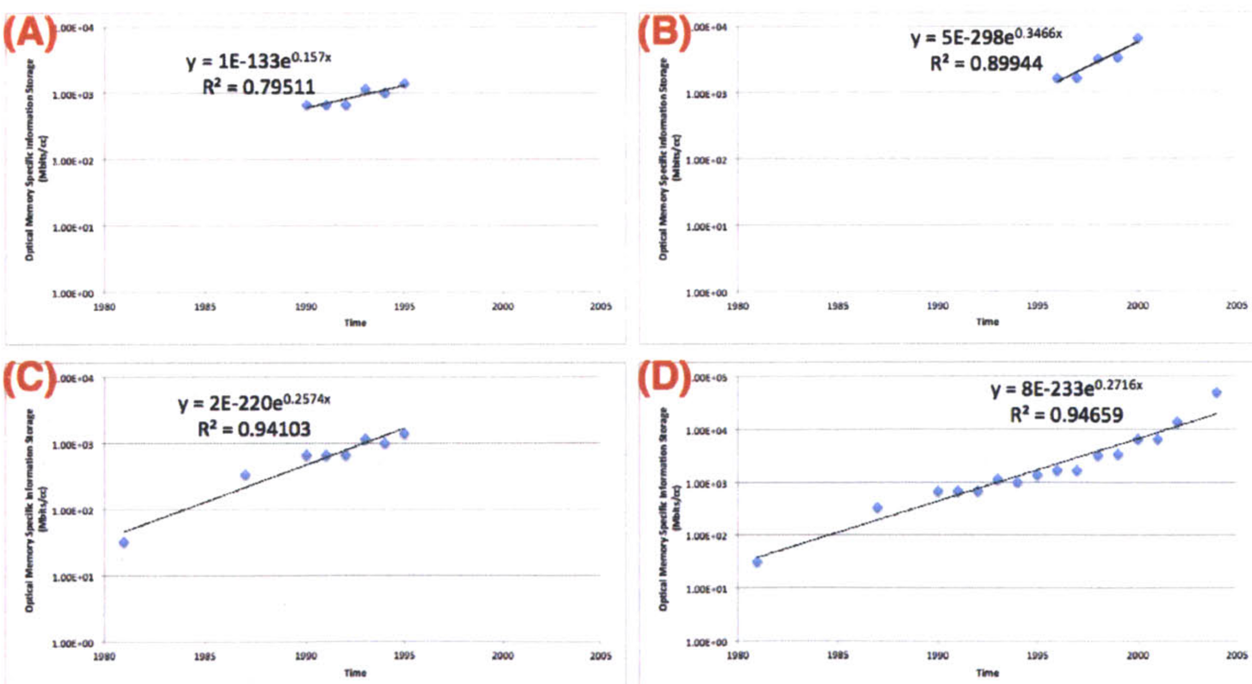



**Figure 26: Optical Information Storage - Mbits/cc (A) 1990-1995, (B) 1996-2000, (C) 1981-1995, (D) 1981-2004: showing that TIRs derived from very short time frames can be unreliable and misleading.**

A reliable TIR should be able to project into the future consistently. Figure 26 (A) shows a TIR of 16%, whereas (B) shows a TIR of 35%, yet both show acceptable $R^2$ values and seem to have a good number of points but yet show a significant difference between the two short time frames. Conversely, the entire timeframe in (C) shows a TIR of 27%, which is consistent with the TIR of 1981-1995 of 26%. This demonstrates that a longer time frame is consistent with obtaining a more reliable and stable TIR.

### 3.2.4.4. Exponential Confidence Interval

Beyond simple metrics there are also statistical techniques to determine a confidence interval for the TIR given a particular set of data points. This can be done relatively easily by the LOGEST function in excel, which provides the standard error for each of the variables in the regression in the form $v = bm^x$, which can be converted to the familiar TIR form as shown in equation 7 ($y = ce^{kx}$) by taking the exponential of both sides with $y = e^v$, $c = e^b$ and $m = e^k$ which allows for the calculation of k. Additionally, the form gives the standard error for m, which can then be converted to a standard error in the same manner: $se_m = e^{se_k}$. This provides an error estimate for a certain confidence interval. If the distribution of the k-values can be assumed as normal, the standard error can be found for a 68% confidence interval ($\pm\sigma$) and 95% ($\pm 2\sigma$) confidence interval. From these 2 values, $\sigma$ can be calculated using the equation 10.

$$\sigma = \frac{se_{68\%} + se_{95\%}}{3} \quad \text{(Equation 10)}$$

Using this value of $\sigma$ and the assumption that possible values of the TIR are normally distributed, it is possible to find the probability that the TIR lies between a certain range of values using z-scores ( $z=\frac{x-\mu}{\sigma}$ ) and a z-score table[2]. While this provides only an average estimate of the standard deviation of the distribution of the potential values of k, this value $\sigma$ provides a measure of reliability for each TIR that can then be used in further Bayesian statistical comparisons.

### 3.2.4.5. Point Removal Method

The TIRs (k-values) are at the foundation of this research and thus should be explored both statistically and empirically. While the standard deviation of the TIR relies heavily on common statistical practices, throughout the course of this research, a more empirical statistical tool was developed called the point-removal-method (PRM). This method complements the other statistical methods and provides further insight into the nature of the TIRs.

The PRM was developed to understand the empirical effects of missing data in a technological improvement curve and also to explore cases without many data points. Throughout prior tests, there was a noticeable effect of specific data points to some of the quantitative technology trends especially with concentrations of points in multiple areas, thus creating the need for a test to determine the robustness of the TIR to missing data points.

The process for evaluating a specific TIR using the PRM is as follows:

For each Technological Improvement Curve

1. Remove Data Point *i*

2. Record $TIR_i$ (k-value) and $R_i^2$ for the $TIC_i$ which is the technological improvement curve with point *i* removed

3. Add data point *i* back to the data set

4. Repeat steps 1-3 for each i from *i=1* to *i=n* where *n* is the number of data point in the technological improvement curve

5. Construct a table of all $TIR_i$ and $R_i^2$ analyze the results

Figure 27 show graphically how the PRM for SLA 3D printing for the Speed/Layer Thickness FPM.

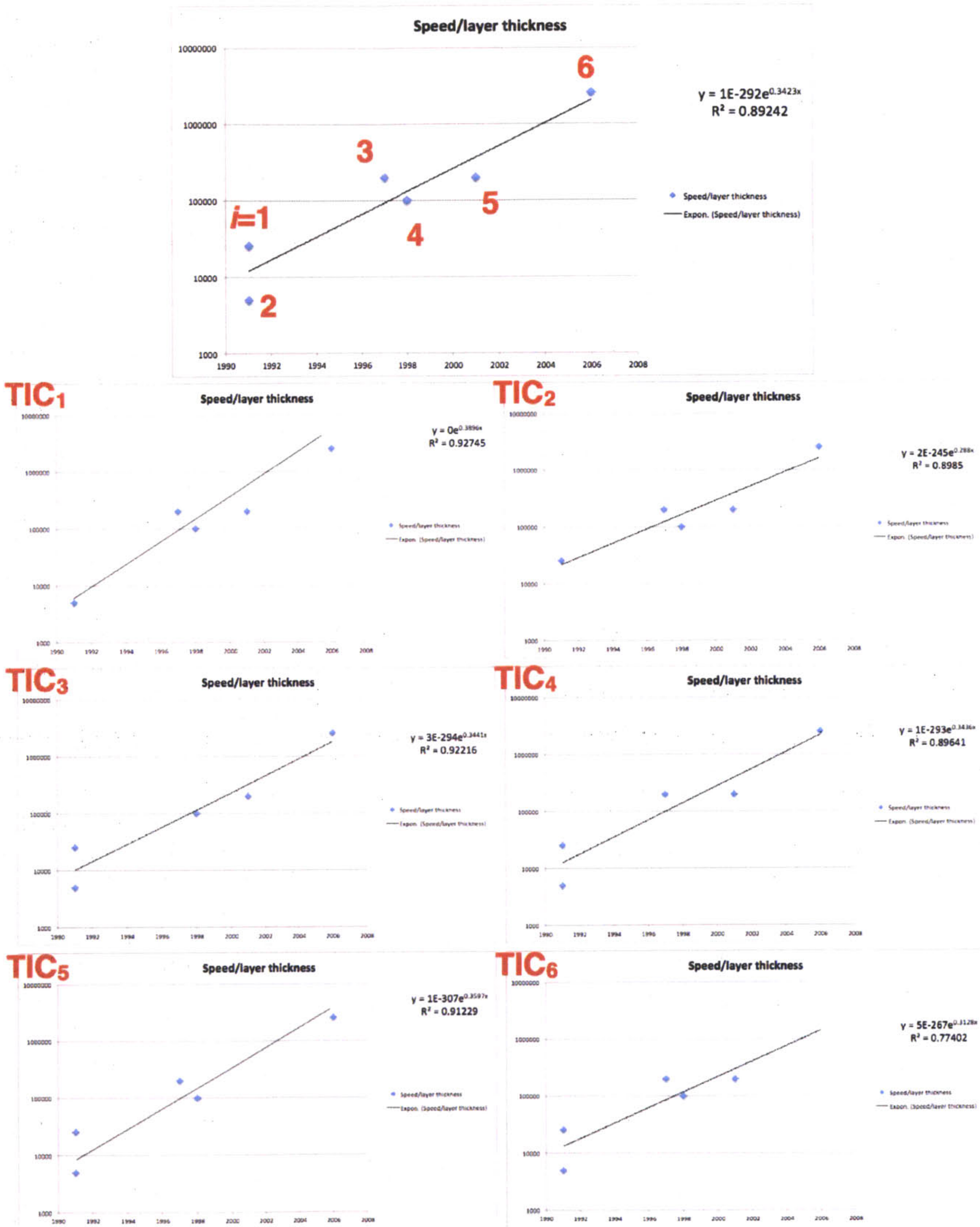



**Figure 27: Graphs demonstrating the PRM method: $TIC_N$ Removes point N from the top graph.**

The first large graph in figure 27 shows the entire technological improvement curve for SLA 3D Printing speed/layer thickness along with the data points labeled i=1 through i=6. Each subsequent graph ($TIC_{1\text{-}6}$) shows the $TIC_i$ with the respect data point removed along with the $TIR_i$ and the $R_i^2$. In this example, despite the relatively small number of data points, the TIR seems relatively consistent. In order to show this quantitatively, the set of $TIR_i$ **and the** $R_i^2$ can be combined into a table as is shown in Table 5.

| Point Removed | K-value | R2 |
|---|---|---|
| None | 0.306 | 0.9 |
| 1991a | 0.319 | 0.899 |
| 1991b | 0.285 | 0.892 |
| 1991c | 0.31 | 0.887 |
| 1997 | 0.303 | 0.93 |
| 1998 | 0.314 | 0.951 |
| 2001 | 0.294 | 0.89 |
| 2006 | 0.32 | 0.828 |
| **STDEV** | **0.012** | 0.036 |
| **MEAN** | **0.306** | 0.897 |
| MIN | 0.285 | 0.828 |
| MAX | 0.32 | 0.951 |
| Range | 0.035 | 0.123 |

**Table 5: PRM values for SLA 3D Printing - speed/layer thickness**

The upper portion of table 5 lists the different k-value ($TIR_i$) and the $R_i^2$ for each of the points that were removed. In the lower portion of the table, some simple statistical measures have been created for the new set of $TIC_i$. Through these statistics it is possible to develop a mean ($\mu_{TIR}$) and a standard deviation ($\sigma_{TIR}$) that can be used alongside an assumption of normal distribution to determine the probability that the true TIR is between certain values. In a more conservative methodology it is possible to use the MIN and MAX value of the TIR to create a bounds for a particular DMP.

### 3.2.5. The most complete and reliable TIR for Each TD

The result of many statistical tests is an understanding of the reliability of the TIR for each DMP. For the cross-domain comparison of patent markers and TIRs, one TIR for each domain was selected for the main analysis on the basis of several different criteria:

- Reliability (measures discussed in previous section)
- Completeness
- Dominated vs Non-Dominated Points
- Date Range

#### 3.2.5.1. Completeness

Completeness of a technological improvement rate (TIR) is an estimate of how well the associated functional performance metric (FPM) represents the purchasing decision of the consumer of the artifact representing the technological domain (TD). This is often the FPM that has the most parameters or the least number of omitted variables. For example, in the manufacturing domains, there are sometimes 4 or more FPMs per TD, and an example of a complete metric would be (speed*build volume)/(layer thickness*price*machine size), which would be preferable to the simpler metric of speed/layer thickness. In some cases where there are multiple FPMs for a TD that have the same number of parameters, the preference is give to the domain without the price parameter (ex: kWhr/kg would be considered more reliable even if less complete than kWhr/$) because of the instability of price-related metrics.

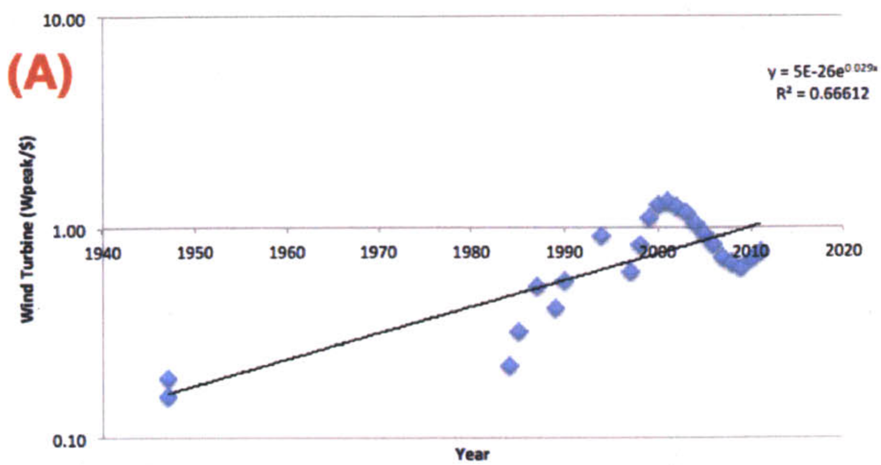
### 3.2.5.2. Dominated vs Non-Dominated Points

When considering how technologies improve over time, there are two ways of including the data points in a technological improvement curve. In most cases, it is desirable to only take into account the 'record-setting' data points. In order to do this, only the <u>non-dominated</u> points are considered, meaning that data points that occur at a later date and have a lower FPM value are removed from the data set. In doing so, the TIR is measuring the improvement of the best example of a technology over time, and does not take into account inferior cases.

While it may seem like non-dominated points should be used in every case, there are a number of reasons why this is not always the case. In some cases, the number of non-dominated points is very small, and thus removing the dominated points reduces the reliability of a technological improvement curve to an unacceptable level. While lack of quality data is a poor reason to choose one approach over the other, it is important to note the other reason for taking into account dominated data points - omitted variables. As was mentioned previously, it is impossible to take into account every parameter that may impact a purchasing decision for a TD, therefore every FPM is certain to have some form of omitted variables, even if the FPM takes into account *most* of the information that affects a purchasing decision. For example, while the size, speed, fuel efficiency, safety and cost of an automobile are likely the main factors in a purchasing decision, other factors such as the stereo system, heated seats, or other amenities are likely to affect consumers' choices as well. Because of these omitted variables, it is possible that some of the dominated points in a technological improvement curve are actually non-dominated if one were to include all of the omitted variables. This effect was shown graphically in Figure 20 in Section 3.2.1 Constructing the FPM. For this research, non-dominated points are used whenever possible as long as the reliability of the TIR remains high.

#### 3.2.5.3. Aligning Time Frames of the TIRs with the Patent Sets

While a long time frame is more reliable, for this research, due to the fact that a major use of the TIRs is to examine them across domains based upon patent characteristics (which we only obtain starting in 1976), it may be desirable to only include data for the TIRs from the timeframe: 1970-Present. Once again there are factors to take into consideration when using only a portion of the data. In particular, once again, data reliability is the main factor in the decision to only use patents since 1970 or to use all of the data points. If, when the data points are removed from the years prior to 1970, the data reliability is significantly lowered, it often makes sense to include all of the data. While the use of pre-1970 data may not align exactly with the patents that are evaluated, the long-term exponential improvement rates have been shown to be relatively consistent, and thus the difference between the 1970-present and the entire data set is rarely very large. In the case where there is a significant variation in TIR across time frames, a deeper analysis into the causes is required. For example, as shown in Figure 28, when studying wind turbines, most of the data points for the technological improvement curve are from examples after 1980 where subsidies have caused much noise (Benson and Magee, 2014), however there are two data points in the very early stages of the technology in 1947.



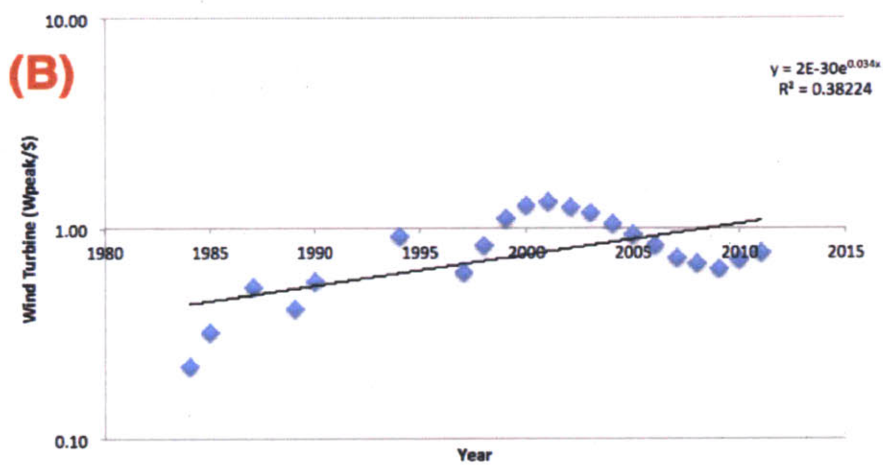




**Figure 28: Comparison of Technological Improvement curves for (A) 1970-present and (B) the entire FPM data set**

The difference between the TIRs is not large (2.9% vs 3.4%), yet the omission of the early data points reduces the $R^2$ from an acceptable 0.67 to a very low 0.38. Upon further analysis, the data from 1970-present is largely affected by factors that are exogenous to technological change, most notable fluctuating subsidies during that time frame, resulting in short-term fluctuations that skew the TIR and make it difficult to understand. It is for reasons such as this that in a few cases, the entire data is used despite the fact that the patents being studied are only from 1976-present.

Ultimately a combination of all of these factors are taken into account to determine the most representative TIR for a given TD, with the ideal case being a TIR that is the most reliable and most complete. There are domains beyond the 28 used in this research that have some data points, but do not have a TIR that is reliable and complete enough to be included in this study.

## 3.3. Manufacturing Performance improvement trends

While the technical performance of many technologies have been quantitatively examined and have been found to follow exponential improvement rates over long periods of time, almost none of these studies have been focused on a manufacturing technical domain. One of the contributions of this thesis is to determine TIRs for four manufacturing domains. 'Traditional' machining technologies have seen the introduction of computer numerical control (CNC) systems, photolithography has improved to the point where billions of transistors can be placed on a single chip, and the introduction of 3D printers has revolutionized the product design process. Because of the importance of these qualitative improvements, it is desirable to establish a more quantitative description of how manufacturing processes have improved.

Most prior in-depth technological analyses of manufacturing processes focus on exploring a specific technology at a specific time. For example, an important paper on nano-manufacturing technologies by Liddle and Gallatin (2011) describes a 'snapshot' of many of the most prominent manufacturing technologies that are capable of producing features on the nano-scale. The paper provides a thorough overview of many different technologies, and describes capabilities of the technologies in an economically significant manner using a comparison of resolution and areal throughput. There have been similar studies completed for metrology (Cullinan et al, 2012) and traditional machining (Kalpakjian and Schmid, 2010). While these studies are essential for understanding the current state of a single technology, a more comprehensive study of several manufacturing technologies across many years is desired in order to understand the relative rates of change of different manufacturing domains.

Other studies have attempted to explore the change in manufacturing technologies over a longer time period through case studies, historical analysis, qualitative analysis and labor productivity metrics (Smith, 1994). These studies often explore the idea of 'productivity advances' as synonymous with the improvement of manufacturing as measured by the productivity level of workers (Jorgenson and Grilliches, 1967). While these studies are essential in understanding the aggregate production function (Solow, 1957), they are usually too broad for analyzing particular manufacturing technological domains. This section of the thesis provides an expansion of the quantitative understanding of changes in manufacturing technical capabilities with time. The intent of this case study is to construct a bridge between the level of detail of the technology specific studies with the long timelines and high level views of the productivity studies as applied to manufacturing domains.

Four manufacturing domains were analyzed that provide breadth and depth in technical, scale and temporal respects. The four domains are milling machines, 3D printers (Industrial stereolithography (SLA)), photolithography machines, and the production of carbon nanotubes.

Determining appropriate FPMS for manufacturing technologies is a difficult task due to the extremely varied output of the technologies. Manufacturing technologies are considered heterogeneous technologies in that they have a variety of functions that may be weighted differently for different customers (Alexander and Mitchell, 1985). For manufacturing, there are four main characteristics of interest: Speed, flexibility, quality and cost. An informed buyer would want to maximize speed and flexibility and minimize cost and dimensional non-repeatability and thus we consider a generic manufacturing metric to be as shown in Equation 11.

$$\frac{SPEED * FLEXIBILITY}{RESOLUTION * COST} \qquad \text{(Equation 11)}$$

In addition to adding to the quantitative understanding of the development of manufacturing technologies over time, this case study will also provide an in-depth look at the selection of domains, the creation of FPMs and the determination of TIR for the four manufacturing domains as an example of the process that is shown for 28 TDs in the results section of the thesis.

### 3.3.1. Milling Machines

Milling machines were selected to provide a baseline for how a well-established manufacturing domain has improved over very long periods of time. Milling machines have been commercially available since the 19th century and provide an interesting comparison to the newer technologies that were selected. A further benefit of studying milling machines is that their long time frame also allows for greater confidence in the results, as it will be less prone to short term data fluctuations than some of the manufacturing domains that have existed for much shorter amounts of time. In addition to the gathering of data for milling machines, we also performed a smaller study into lathes, which returned less data, but allowed for a comparison between two similar manufacturing domains.

#### 3.3.1.1. Milling Machine FPM

There are several measures that can be used to evaluate speed and flexibility of Milling Machines. The most fundamental approach to characterizing the speed of a milling machine is the material removal rate, which is the volume of material removed by the machine per amount of time. The material removal rate can be calculated in three different ways involving differing parameters as shown in equation 12.

$$MRR = v_{workpiece} * w * d$$
$$MRR = f * \Omega * w * d \qquad \text{(Equation 12)}$$
$$MRR = \frac{P}{\mu_s}$$

Where $v_{workpiece}$ is the velocity that the work-piece is moved, $w$ is the width of a particular cut, $d$ is the depth of a particular cut, $f$ is the feed rate in inches/revolution, $\Omega$ is the rotational speed of the bit in revolutions per minute, $P$ is the power of the machine and $\mu_s$ is the specific cutting energy. Several of these parameters are very dependent on the specific cut that is being performed using the machine. Any of the methods that depends on cut-specific parameters such as width and depth of cut are not at all likely to be stable over time and thus not allow study of material removal rate over time. The final method of calculating the material removal rate (shown after the 3rd = sign) takes into account only the power of the machine (to the tool) and the specific cutting energy, which is very closely related to the material properties of whatever is being cut. It is for this reason that we use the power of a milling machine as the appropriate measure for its speed. While power of a machine is our first choice for evaluating speed, we were also able to compare the max rotational speed ($\Omega$) of the machines, which provides a check on the reliability of the power data.

Measuring flexibility can be done by analyzing the size of the part that can be built or by looking at the degrees of freedom a particular milling machine has. Due to the limited nature of the degrees of freedom measure, it is not an appropriate metric because it is necessarily limited to only a few values. The table size is a more appropriate metric due to its non-discrete nature. The accuracy of a milling machine is a given metric for many machines and measures how closely a machine can conform to the demands given by the operator (how close to the bull's-eye it hits), and results in an improved machine when the measure decreases, which is consistent with the general manufacturing metric template. Finally, cost can be evaluated using inflation-adjusted dollars and the machine also improves as the measure decreases.

#### 3.3.1.2. Milling Machine Data Collection

One challenge in constructing a robust quantitative performance trend for milling machines was finding samples of data from a long time period. The source we found most useful was the trade journal *Modern Machine Shop*, which has been published continuously since 1928. This magazine includes articles about improvements in machining tools, as well as advertisements for particular machine tools that include many specifications that can be used to compile the metrics described above. Figure 29 shows an example of an advertisement from an early issue of *Modern Machine Shop*. This advertisement contains a number of different specifications such as spindle speed (125 to 1600 revolutions per minute), feed rates (0.004 to 0.025 in/revolution) and drill size (1/2 to 1.5 inches in diameter). A combination of these metrics was used to construct the technical performance trend and allowed calculation of the TIR.

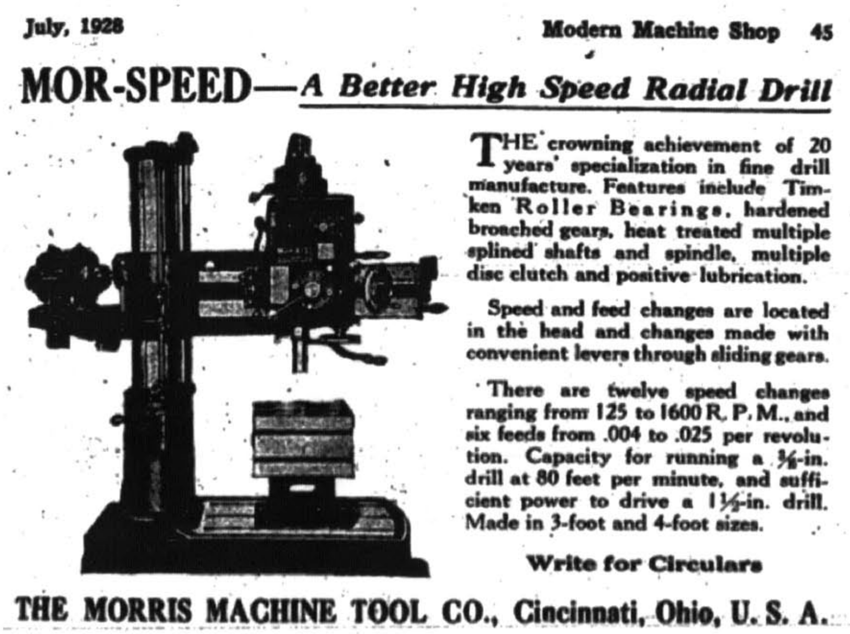



**Figure 29: Example advertisement from Modern Machine Shop that gives performance values in the final paragraph – adapted from (Modern Machine Shop, 1928).**

Data points from 1928 until the present day were obtained from this source. A drawback of gathering data from Modern Machine Shop is that each advertisement or article in the magazine does not include all of the specifications needed to form each metric. The result of this is a rather sparse matrix that includes many different machines from a wide time frame, each with a different set of specifications. In order to process this data, we developed a data reduction method that uses the yearly averages as well as the individual machine metrics for many of the specifications.

This process, named the data averaging yearly collection (DAYC) method, involved taking each of the metrics and finding the mean of the specification for a given year. For example: if there were 15 milling machine data points in 1964 and some of them had power ratings and others had accuracy specifications, we take the average of the power ratings and the average of the accuracy specs for that particular year and create a data point that represents the average speed/quality of milling machines for 1964. This data is then combined with the data points from the milling machines that included both data points. Figure 30 shows an example plot with this process applied to the max table speed of a milling machine, it shows consistency between data obtained by the yearly averaging method and the data from individual machines.

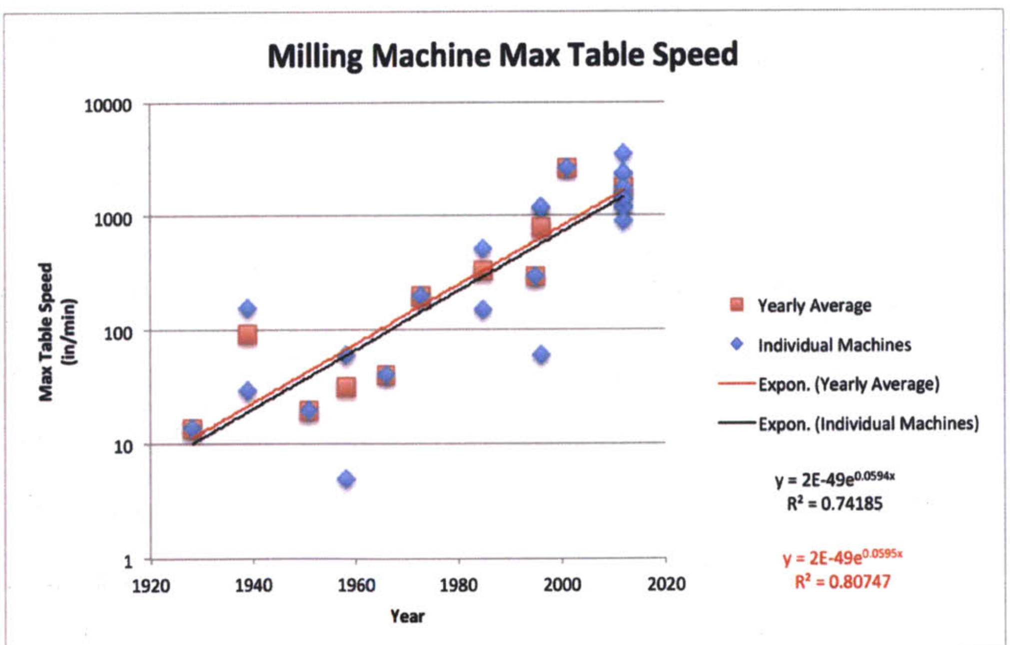



**Figure 30: Example of data averaging yearly collection (DAYC) method**

The DAYC method also provides an additional benefit of reducing the noise in the specialty machines that are made specifically to maximize one of the constraints without concern for the others. An example of this would be a very large and powerful milling machine that was specifically designed for rough cuts, and therefore sacrificed accuracy for speed. These specialty machines are an interesting showcase of the ability to stretch the measures at a certain point in time, but are not representative of the general improvement rate of a particular technology.

We were able to find data for 100 milling machines from July 1928 until August 2012, each with various specifications relating to their speed, quality, flexibility or cost. Our highest confidence metrics are Table Speed/Accuracy and Machine Power/Accuracy, and both were compiled using the DAYC method.

#### 3.3.1.3. Milling machine technical capability curves

Two reliable technological improvement curves were completed for the milling machine TD. Both of the metrics take the form of speed/quality. As mentioned above the most complete metric available by the data is the comparison of machine power to represent speed with the repeatability of the machine to represent quality. Figure 31 shows the improvement rate of milling machines power/accuracy since 1928.

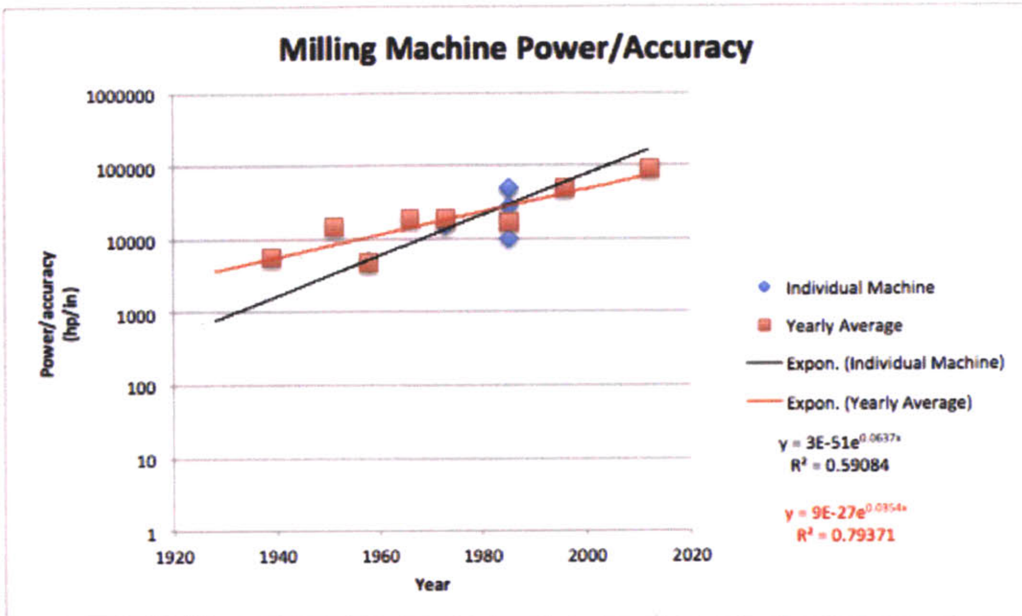



**Figure 31: Improvement rate of milling machines measured using Power/Accuracy**

The graph shows an improvement rate of 3.5% ($R^2 = 0.8$) for the DAYC method and an improvement rate of 6.4% ($R^2 = 0.6$) for the sparse individual machine data. The improvement rates differ slightly between the two data collection methods, with more confidence in the DAYC method's results due to the larger number of data points and the higher $R^2$ value. However, we also note that the individual machine data points when combined with the DAYC lead to little change to the DAYC points alone ($y = 3.84$ and $R^2 = .67$)

The other metric that was collected was milling machine table speed/accuracy and is shown in Figure 32.

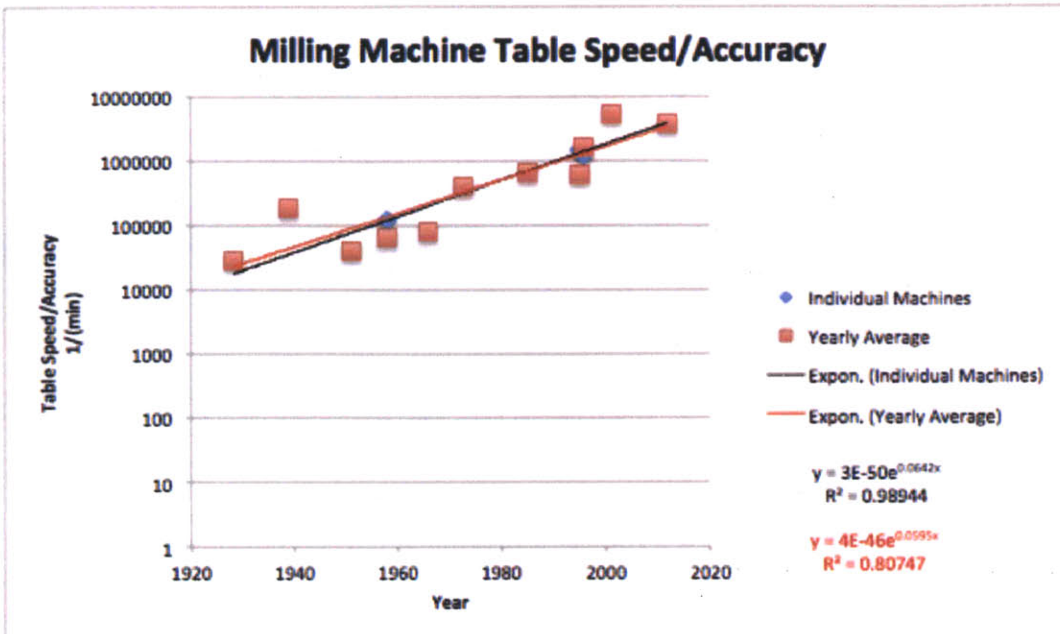



**Figure 32: Improvement rate of milling machines measured using Table Speed/Accuracy**

The improvement rate of this metric shows a k of 6% for the DAYC method ($R^2 = 0.81$) and 6.4% for the individual machines ($R^2 = 0.99$). While this metric is not a perfect representation as mentioned earlier, it does corroborate the improvement rate seen in the more representative metric of machine power/accuracy. When looking to analyze milling machine speed/quality incorporating all of the metrics, the domain improves at approximately 5±1.5% per year. This improvement does not include cost measures for milling machines at this point, which is an area of strong interest for future research.

### 3.3.2. SLA 3D Printing

There has been considerable interest in 3D printers in recent years, despite the fact that the first commercially available 3D printer was introduced to the market nearly 20 years ago in 1987 (3D Systems SLA-1). Because the term '3D printer' has been used to describe a number of different technological domains, it is necessary to be more specific about what technology is being studied. Partly for data availability and length of time on the market, 'industrial SLA 3D printers' was selected as the domain to represent the field of 3D printing/additive manufacturing. In particular, the focus is on industrial 3D printers, as opposed to hobbyist 3D printers because of their very different levels of maturity and the hobbyist market is much more recent. Many of the hobbyist 3D printers carry significantly less capability than their industrial counterparts in many of the areas that are important in 3d printing such as flexibility, and reliability of produced parts. The stereolithography (SLA) technology was selected in particular because it has the longest time frame of commercially available machines. Other technologies such as fused deposition modeling (FDM), selective laser sintering (SLS) and the inkjet/powder-based technologies might also be studied as well, but offer less time for establishing data trends.

#### 3.3.2.1. SLA 3D Printing FPM

For 3D printing, there are also a number of different ways to evaluate each of the core capabilities of the machines. In order to understand the particular capabilities of how the machines perform, it is important to understand the underlying mechanisms behind the technology being studied. In the case of stereolithography, the parts are produced when a laser cures a pattern on a thin layer of polymer, then a new layer of polymer is added and the laser cures another pattern on the next layer. This process is repeated many times to form an object that is suspended in a bath of liquid resin. The part can then be removed, cleaned and post-processed if necessary. Figure 33 shows a diagram of how the process works.

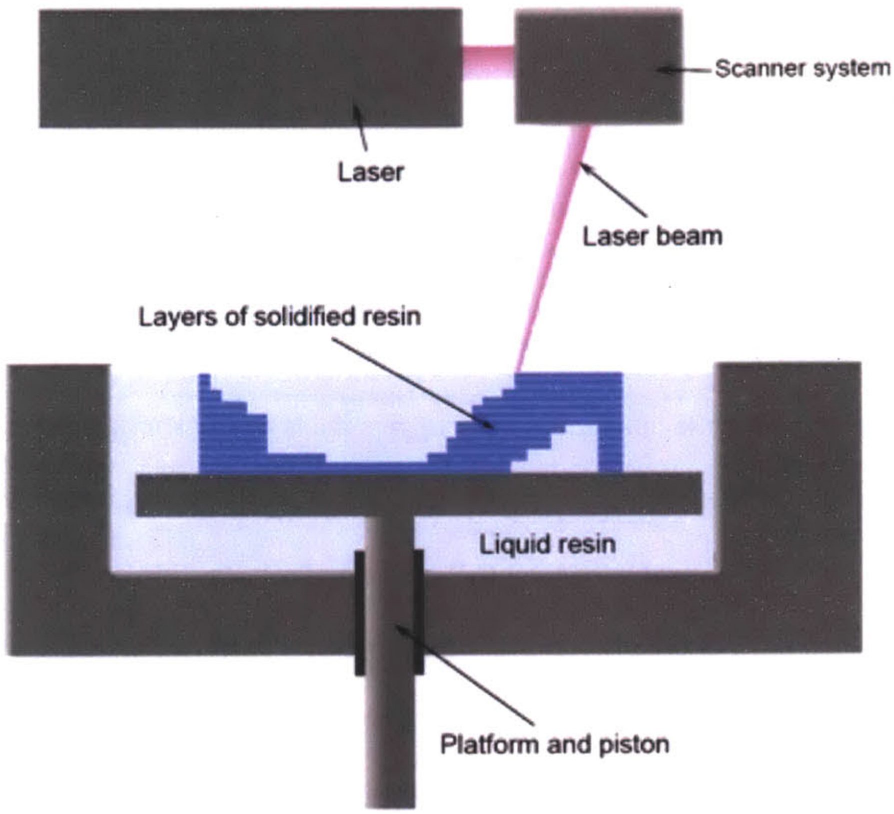



**Figure 33: Graphical Depiction of the SLA 3D printing process**

In order to measure speed of an SLA printer, one must consider both the amount of time it takes for the laser to scan one layer of resin, the time it takes to add another layer of resin, and the post processing time. Ideally data would exist for each of these different times for each machine, however the industry has adopted a standard of only giving the speed that the laser moves, which can be related to the amount of time it takes to scan and cure one layer in the SLA-produced object. In this research, the measure of speed for the industrial SLA 3D printers used is the laser speed measured in mm/sec. The use of this measure for speed is not perfect, as it does not take into account other speed-based considerations such as the time it takes to move from layer to layer (which is often considerable) or the time for post-processing (which is usually much less than the print time), nonetheless the measure of laser-speed does provide a usable metric for the speed of an SLA 3D printer.

Like the measure of speed, there are a number of factors that can contribute to the quality of a 3D printed part. The list can be quite extensive, including measures such as layer thickness, repeatability and material properties of the parts being produced. The industry standard for determining the quality has been to give the layer thickness, which is simply a measure of the thickness of each of the particular layer of resin that is being cured at one time. This measure is also interesting in the fact that a smaller layer thickness will result in a longer build time for each part, as the process will require more steps of curing and adding layers. The selection of layer thickness as a measure of quality complements the selection of laser speed for the rate of the manufacturing process because they are largely independent of each other.

Some of the other metrics that could be used for quality include the repeatability or accuracy of the machines, and the material properties could be a useful measure of flexibility. While these are important metrics for a SLA printed part, they are not widely recorded for each of the machines and thus are not a reliable metric. It is possible that the industry trends may change and different metrics will be tracked, allowing us the ability to understand how these other measures of quality have changed over time, but at this time the data proves to be too sparse for a reliable technical capability metric.

One of the main benefits of 3D printing technologies is the flexibility that the processes provide for the engineers and designers. Measuring the flexibility of manufacturing processes has been studied at length, and continues to be a topic of much discussion. For a 3D printer, the main metric that allows for measurement of flexibility is the print build volume, which is the size of the largest part that the machine can produce. While there are other possible aspects of flexibility of a 3D printer, including the number of different materials that can be used and the robustness to many designs (such as items with overhangs), the overall print volume remains the most widely reported and is a more objective and continuous metric than some of the others.

The cost of an industrial SLA 3D printer can take a number of forms. The most obvious measure of cost for a 3D printer is the inflation-adjusted price in US dollars. This metric is used to evaluate cost for this domain. In addition to the price of a machine, there is also a cost in the amount of space required for the machine to be installed. This trend can be seen in the movement from large appliance-sized 3D printers to many of the desktop 3D printers that are seen today. The reduction in size of the machines allows for them to placed in smaller areas and can reduce the costs of moving such a large object and thus can reduce the costs associated with the machine.

Overall there are five measures that we can combine in different ways to form metrics to measure the increasing technical capabilities of industrial SLA 3D printers. Equation 13 shows a metric including all five of the measures, which can be manipulated in a number of different ways to explore the technical dynamics of the domain.

$$\frac{LaserSpeed * BuildVolume}{LayerThickness * Cost * MachineSize} \qquad \text{(Equation 13)}$$

For example, one way to manipulate these measures is to remove the cost and flexibility metrics and simply analyze the speed/quality metric to understand how the core capabilities of the machine have improved over time. In adding or removing measures from this metric and seeing how the improvement rates differ, one can learn a significant amount about how the domain has improved over time.

#### 3.3.2.2. SLA 3D Printing Data Collection

While 3D printing has gained more attention in recent years, its roots can be traced back to the early 1980s when a Japanese researcher published his work on an early SLA machine (Kodama, 1981). While the first commercially available industrial SLA 3D printer was not available until Charles Hull and 3D systems released the SLA-1 in the late 1980s. In the early 1990s, the first widely available 3D printers were released by 3D systems call the SLA-190, SLA-250 and SLA-500. Our data combined a number of different sources, and contains data as far back as 1991 when 3D-systems released their trio of printers. While some of the early data from these machines comes from academic journal articles (Kruth, 1991; Waterman, 1994), most of the data for the industrial SLA printers was gathered from a yearly review publication called the Wohler's Report (Wohlers, 1996-2010) and the complementing specification sheets from the manufacturers.

Compared with the data for milling machines, the data for industrial SLA 3D printers has fewer, yet more complete points. There are 19 industrial SLA 3D printers in our collection of data over the last 20+ years and there are 7 with a complete set of specifications for the metric described in equation 13 and the remaining 12 are only missing one or two out of the five measures that make up the most complete metric described above. For example, all 19 of the data points include cost, build volume and machine size, with 12 of those also including layer thickness, and 7 of those also have laser speed. The completeness of this data allows for more reliable analysis of the technical improvements of the domain. Ultimately we were able to compile improvement curves for 4 metric-domain pairs ranging from laser speed/layer thickness to the most complete metric described in equation 11.

### 3.3.2.3. SLA 3D Printing technological improvement curve

There was significant doubt among those that we spoke to about the ability to create a technical capability curve for 3D printing. Part of the reason is because the individual measures do not seem to follow any sort of pattern, and thus are difficult to use to find an improvement rate for the domains in the 3D printing area. For example, one metric to consider would be the flexibility of the industrial SLA 3D printers, which can be measured by the build volume. Figure 34 shows the change in build volumes of industrial SLA 3D printers since 1991. The data does not indicate any sort of temporal relationship or consistent improvement in this individual measurement.

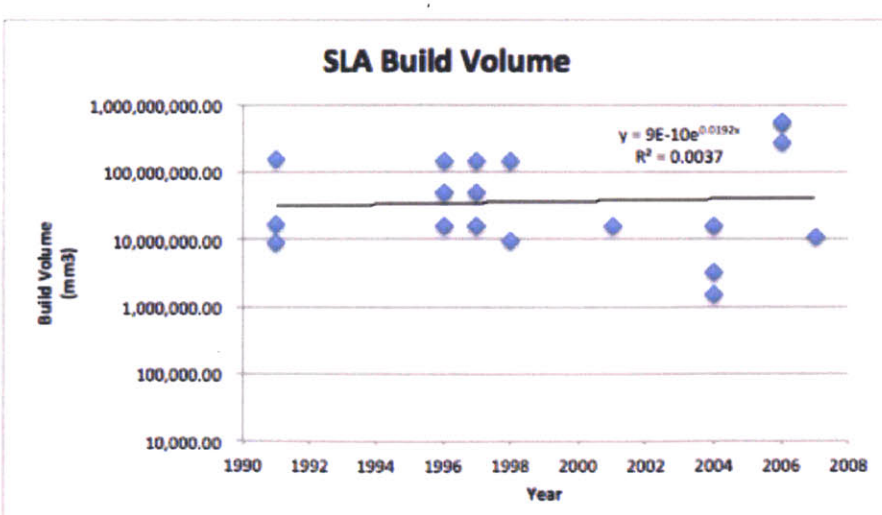



**Figure 34: Logarithm of SLA Build Volume over time**

While the individual measurements of industrial SLA 3D printers do not create a clear picture about the improvement rates of the domain, the more complete FPMs discussed above shows a more convincing relationship between the improvement of SLA 3D printers and time. Figure 35 shows the improvement of laser speed/layer thickness for the domain.

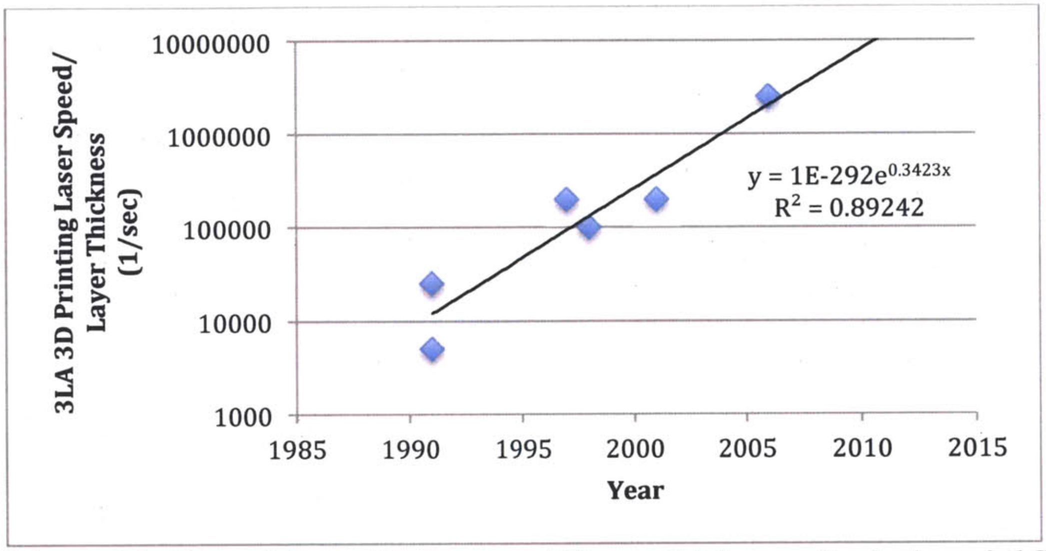



**Figure 35: Logarithm of laser speed/layer thickness for industrial SLA 3D printers**

The data shows an improvement rate of 34% for the technical metric with a very strong $R^2$ of 0.89. Note that all of graphs for this domain show only the individual points and not the DAYC method due to the smaller number of available points and the more complete data for each of the points. Figure 36 shows the improvement curve of the technical metric when the price of the machine is also taken into account.

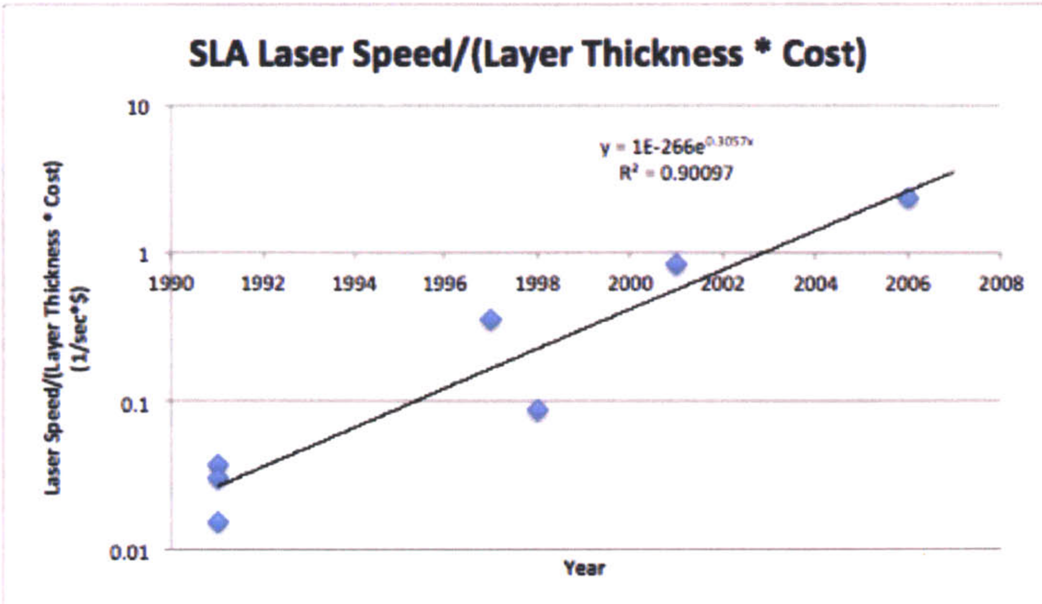



**Figure 36: Logarithm of laser speed/(layer thickness*cost) vs time for industrial SLA 3D printers**

When the economics are taken into account, the improvement rate of the domain is 31% with a healthy $R^2$ of 0.9. Figure 37 adds in a measure of flexibility for the domain.

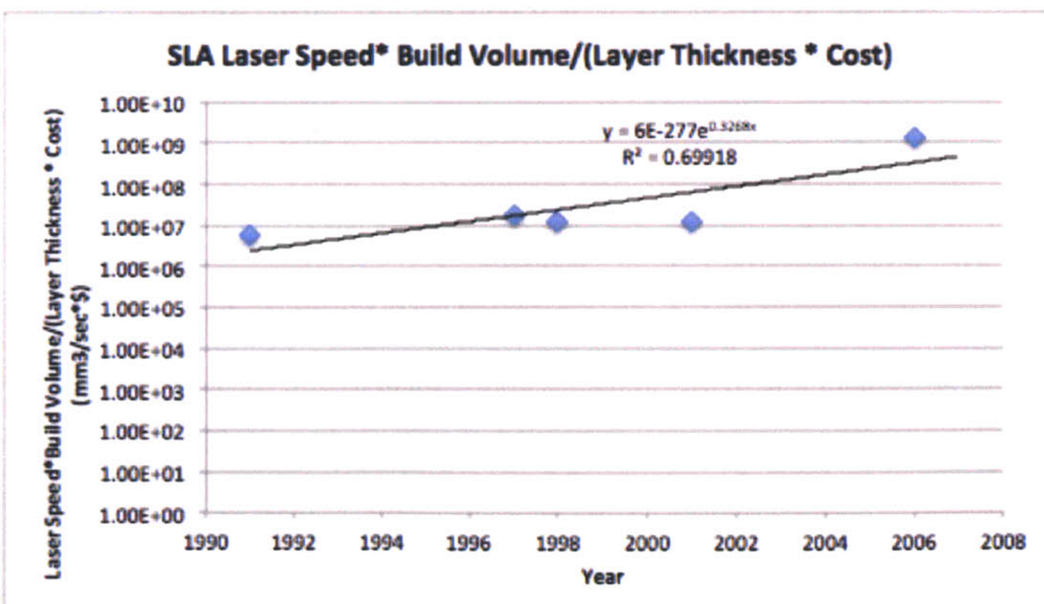



**Figure 37: Logarithm of (laser speed*build volume)/(layer thickness*cost) vs time for industrial SLA 3D printers**

When the flexibility factor is taken into account, the improvement rate remains close to the previous metrics at 32.7% with a reduced $R^2$ of 0.7. The final and most complete metric that was mentioned in Equation 13 is shown in Figure 38.

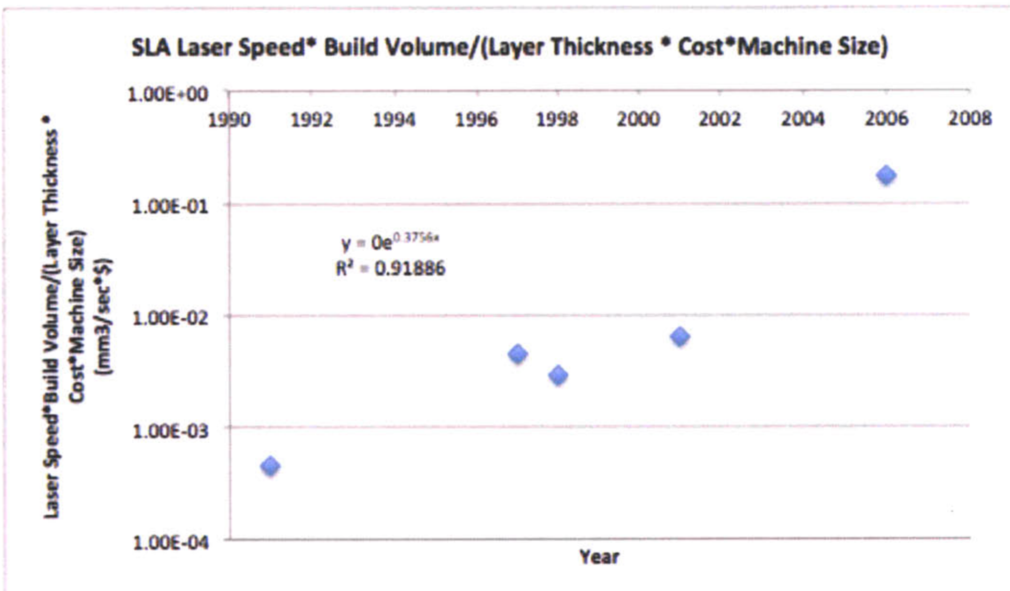



**Figure 38: Logarithm of (laser speed*build volume)/(layer thickness*cost*machine size) vs time for industrial SLA 3D printers**

The improvement rate of the most complete metric is 37.6% with an impressive $R^2$ of 0.92. When all of the different metrics are considered for the domain of industrial SLA 3D printers, the improvement rate of the domain is approximately 31±3% per year and is basically the same with cost in the metric.

### 3.3.3. Photolithography

Due to the extensive literature published concerning Moore's law, an interesting question involves understanding the underlying mechanisms contributing to the improvement in transistor density. It is for this reason that photolithography was selected as a third manufacturing domain. Photolithography is an essential step in the production of microchips, and in its most advanced state is used for relatively few other purposes, making the link between the improvement in photolithography and microchips very strong. A goal of this research is to understand how the improvements in photolithography have compared with Moore's Law and the improvements in microchips. Additionally, photolithography provides a medium length time frame for analysis – as the photolithography industry has existed for over 50 years (less than that of milling machines but more than that of 3D printing).

#### 3.3.3.1. Photolithography FPM

The domain of photolithography follows the same generic formula that the previous two manufacturing domains used. The speed and quality measures are once again more complex than the cost measure, and in photolithography we did not find a reliable metric for evaluating the flexibility of this future-creating technology. A photolithography machine (call a stepper) works by shining a wavelength of light (often in the UV spectrum) on a mask through a lens and onto a wafer (usually silicon). The mask creates a pattern that can be repeated many times on the wafer and is amplified using the lens as is shown in Figure 39.

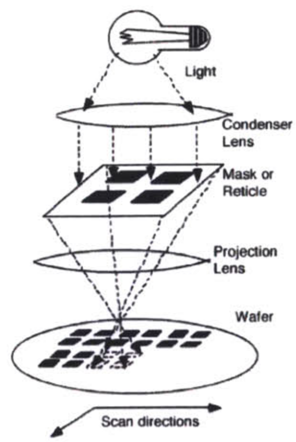



**Figure 39: Mechanics of a photolithography stepper – adapted from www.cnx.org.**

One of the most common measures of speed for a photolithography machine is the throughput of the machine. The main metric for speed is given as the throughput measured in wafers/min. While this is an acceptable metric – it doesn't take into account the increasing wafer sizes that have taken place over the years. One silicon wafer can make many microprocessors and thus over the years the size of each silicon wafer has increased to allow for each wafer to make more microchips. Figure 40 shows the progression of the different sizes of wafers that have been used in the microprocessor industry over the last 50 years. The change in wafer size has resulted in a skewed metric for throughput, as a smaller throughput value may actually produce more microprocessors because of a larger wafer size. Thus a more appropriate metric for speed of photolithography is the areal throughput, which is calculated by multiplying the throughput of the stepper by the area of the wafer that is being used to give a metric of $mm^2$/min.

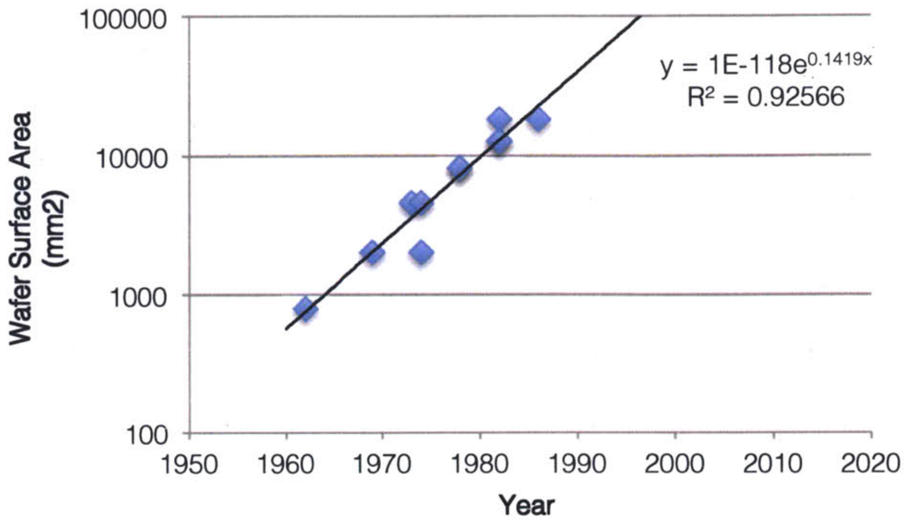



**Figure 40: Different sizes of silicon wafers for use in photolithography steppers over time**

To analyze photolithography quality, there are a number of different measures to look at. The most direct measure is the stated resolution of the photolithography stepper – which describes how small a feature that it can produce and is often measured in microns. A closely related measure is the feature size or the gate length of the integrated circuits that are produced by the stepper. While these metrics are related to the product being produced and are not necessarily specifications of the machine, they are very closely related to the stepper and thus can reliably act a measure of the quality that a photolithography stepper can produce. The minimum feature size and gate length of integrated circuits are also measured in microns and the photolithography system improves as the value of all three quality measures decreases. Ultimately we were able to use all three metrics together to provide for a very complete picture of how the quality of photolithography has improved over time.

Throughout our research, we did not find a consistently reliable metric for flexibility of a photolithographic process. In some ways the flexibility of the process is related to the quality as described above, and thus we will assume that the quality metric takes into account some measures of flexibility as well. Finally, the cost of a photolithography stepper is measured by the standard inflation-adjusted price of the steppers in US dollars. When these measures are combined into a metric for photolithography as shown in equation 14, they can provide a relatively complete picture of how the technology has changed over time.

$$\frac{\mathit{Throughput} * \mathit{WaferSize}}{\mathit{Accuracy} * \mathit{Cost}} \qquad \text{(Equation 14)}$$

Like the most complete metric used for 3D printing, the measures in this metric can be combined in different ways to gain a more complete picture about how the technology has improved over time.

### 3.3.3.2. Photolithography Data

There is a rather complete set of data available for photolithography, both on the specific steppers and on the costs in the industry as a whole. The data for the individual photolithography steppers was collected by a Harvard University thesis of Rebecca Henderson (1988). The thesis includes 11 data points for individual photolithography steppers spanning from 1962 until 1986. The data in these reports was very complete and included costs, throughput, wafer sizes, and resolution.

Gordon Moore provided the yearly average data for the semiconductor industry in his reflective paper on Moore's law that covered the years 1960-2005 (Moore, 2006) as is discussed in depth in section 2.1 of this thesis. His paper includes many different graphs ranging from the cost per transistor in a computer processor to the average cost of lithography equipment for the semiconductor industry. Additionally, Moore gives values for two different measures of quality of the photolithography process: gate length and minimum feature size.

The two sources complement each other in that they provide both a specific stepper point of view along with the industry average higher-level perspective. The confidence in the reliability of these data is increased by their close correlation with each other. The result is four different metric-domain pairs including the most complete metric as described in equation 12.

#### 3.3.3.3. Photolithography technological improvement curve

The photolithography technical capability curves have a large amount of complete data and therefore should be seen as more reliable than some of the other metric-domain pairs, as is discussed in section 3.2.4 of the thesis. Due to the large amount of data we were able to make a large number of reliable curves for individual measures as well as the more telling and more complete metrics related to equation 12.

As was mentioned above, there were 3 measures that could be used to measure quality of the photolithography domain. Figure 41 shows the comparison between the three different measures of accuracy for photolithography along with their inverted exponential regressions.

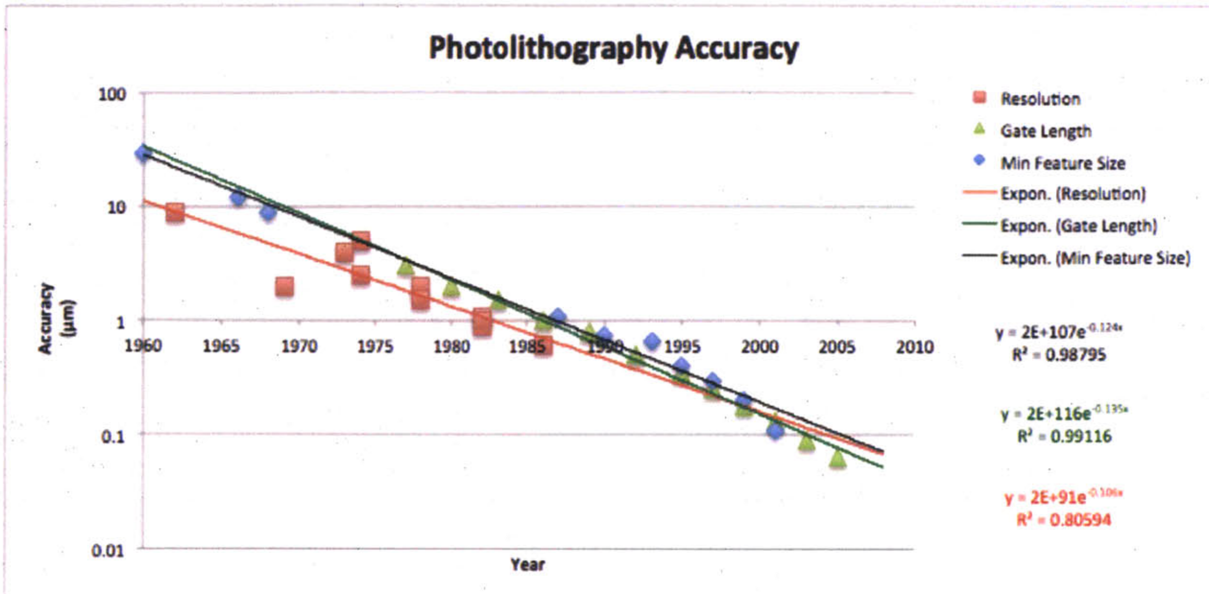



**Figure 41: 3 Different ways of measuring accuracy of photolithography: Resolution, Gate Length and Minimum Feature size**

The data shows very similar improvement rates of 10.6% ($R^2$ of 0.81) for stepper resolution, 13.5% ($R^2$ of 0.99) for the gate length of an integrated circuit and 12.4% ($R^2$ of 0.99) for the minimum feature size of an integrated circuit. The similarity of these measures made it possible to combine the measures into one accuracy measure using all of the data points. In addition, because the technology is improved when the accuracy value decreases, we plot 1/accuracy as shown in figure 42.

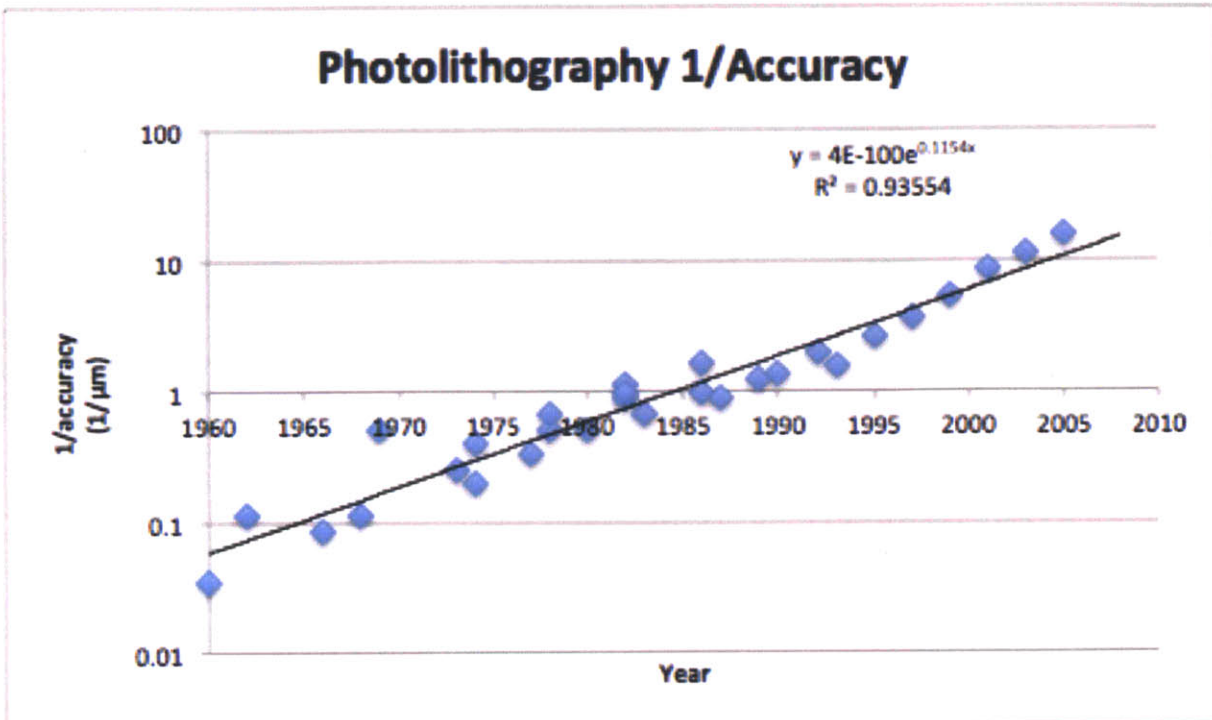



**Figure 42: Logarithm Combined accuracy measure for photolithography vs time**

The aggregate improvement rate of the accuracy of photolithography steppers is 11.2% with an $R^2$ of 0.94. Another individual measure that was given is the photolithography tool cost. Figure 43 shows the increase in photolithography tool costs over the last 50+ years.

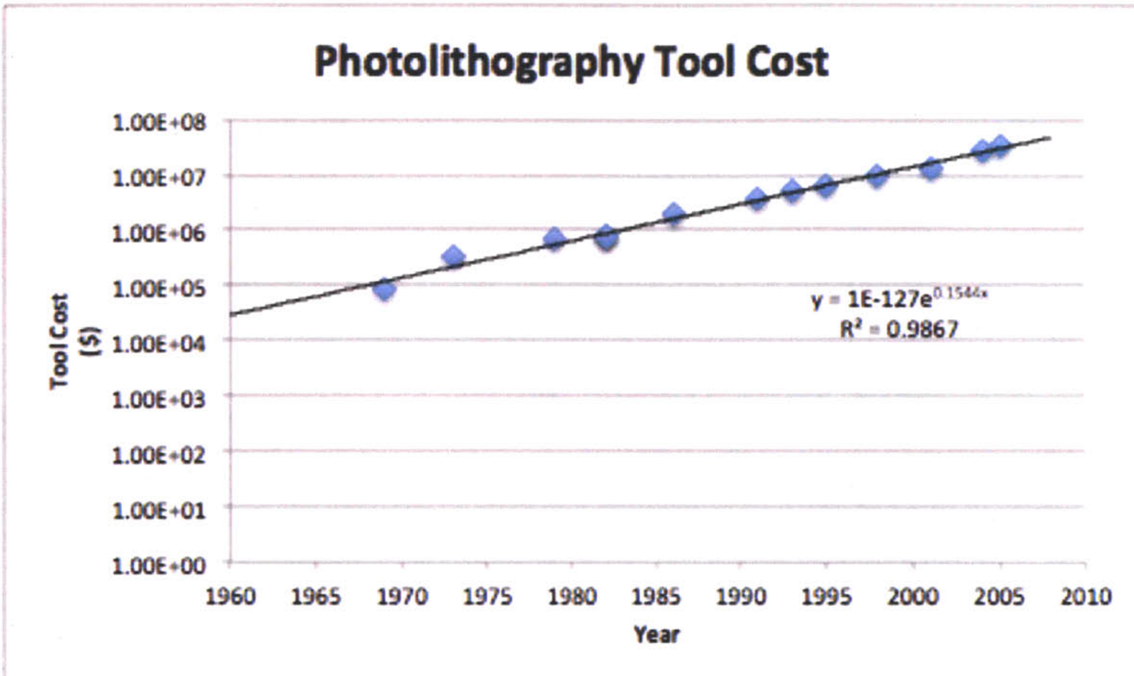



**Figure 43: Logarithm of Tool Price (in 2006 $) vs time for photolithography**

The annual inflation-adjusted cost increase for photolithography steppers is very high at 15.4% ($R^2$ =0.99). Figure 44 shows the technical metric that includes speed and quality of the photolithography steppers.

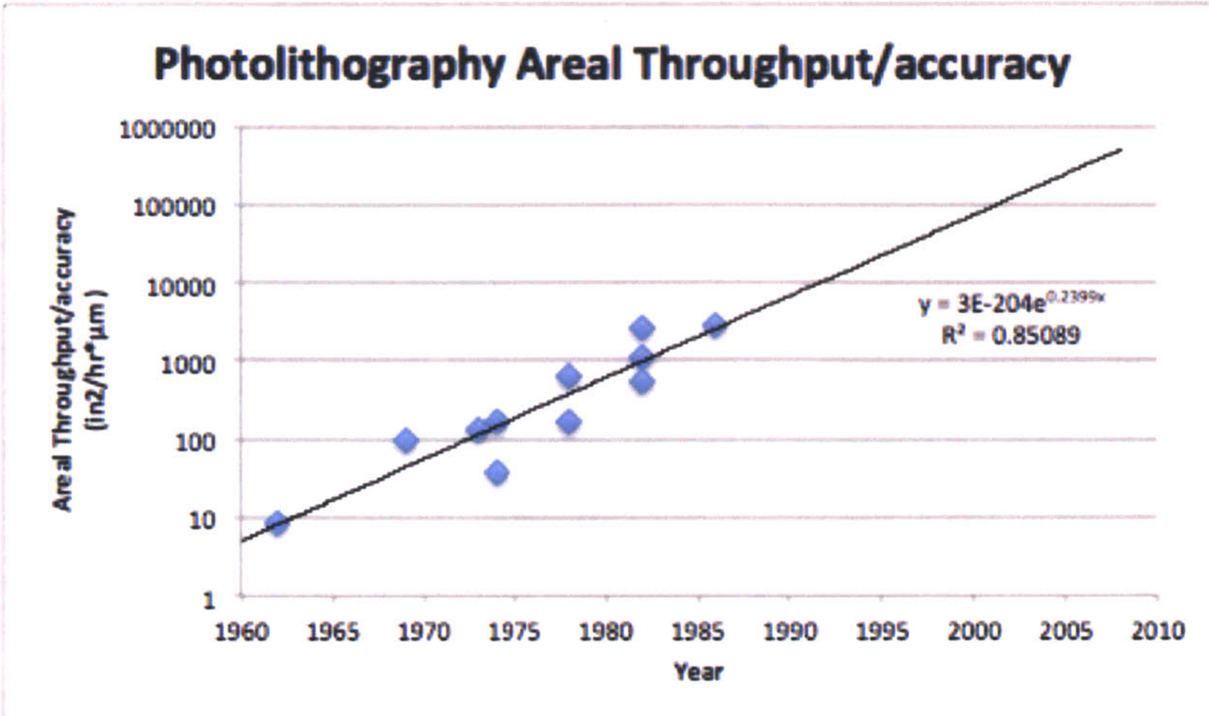



**Figure 44: Logarithm of photolithography areal throughput/accuracy vs time**

The improvement of areal throughput/accuracy or photolithography steppers occurs at a rate of 24% per year with an $R^2$ of 0.85. When the significantly increasing costs are included in the metric, the improvement rate is lower, as is shown in Figure 45.

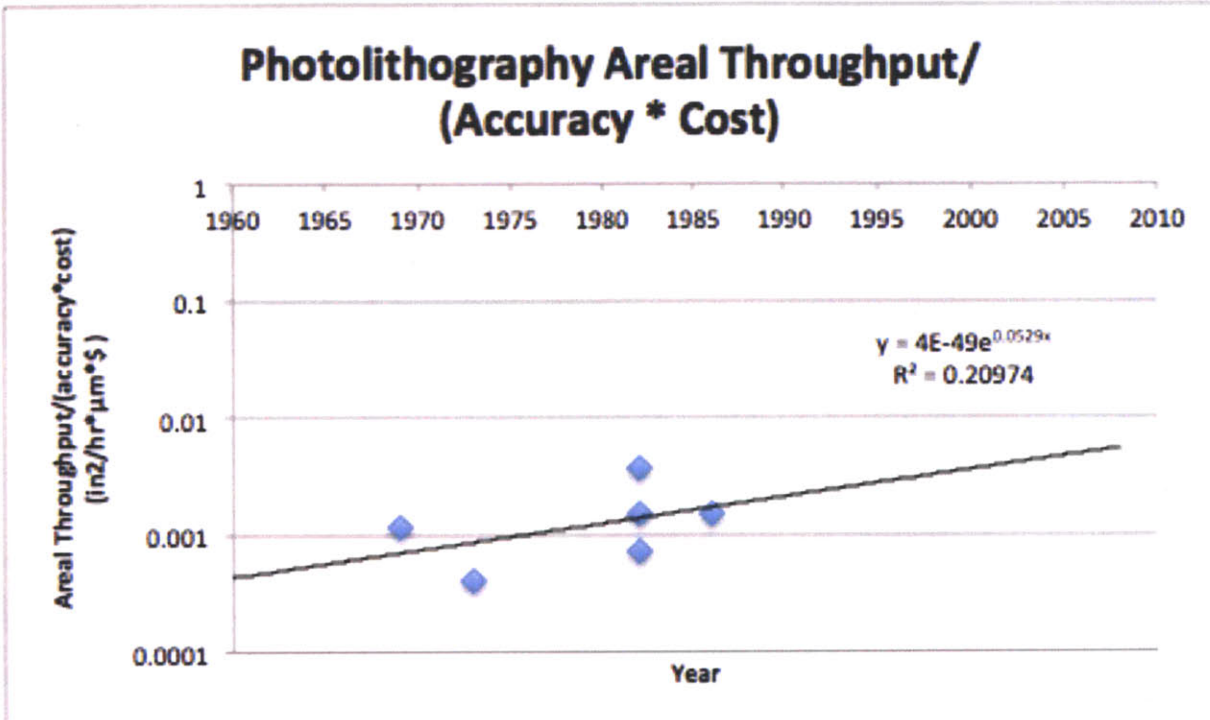



**Figure 45: Logarithm of photolithography areal throughput/(accuracy*cost) vs time**

While the $R^2$ of the most complete metric is very low (0.21), we are more confident in this particular metric due to the high $R^2$ of the relating measures. Areal throughput/accuracy has an $R^2$ of 0.85 and the cost measure has an $R^2$ of 0.99, and the difference between those two improvement rates (24%-15.4% = 8.6%) is around the same value as this particular metric. The combination of all of these factors indicates that the complete metric has an improvement rate of 7±1.6%.

### 3.3.4. Carbon Nanotube Production

Finally, there exists a need to understand how manufacturing at the nano-scale has improved and thus the production of carbon nanotubes (CNTs) is studied. Carbon nanotubes are arrangements of carbon atoms with incredible large aspect ratios that create unique materials properties such as high electrical conductivity, hydrophicity, and strength to weight ratio. There are two types of CNTs: multi-walled, which are produced in large quantities and are relatively less expensive and therefore are the more commonly used CNT; and single wall CNTs, which are more expensive to produce and whose material properties can be more precisely fine tuned to create specific material properties. As with 3D printing, there are several technological processes (domains) used in the production of carbon nanotubes, and while this particular domain is not a strictly manufacturing domain, it should provide a heuristic to how the capabilities of how the combination of chemical vapor deposition and other nano-scale technologies have improved over time. CNTs are a bulk nanomaterial therefore representing a chemical process relevant to nanotechnology, and a raw material that is fed into many products/processes. This domain is also interesting in that it provides a chance to study a nascent industry that is still forming and provides a view of a technology with a very short time frame.

#### 3.3.4.1. Carbon Nanotube Production FPM

Finding reliable metrics for understanding how the nascent field of carbon nanotube production has improved has proven to be a more difficult task than that of the three previous domains discussed in this paper. The relative youth of the CNT industry makes it difficult to gather data of commercially available nano-fabrication methods. The inherent complexity of the processes also contrasts the unit machining operations of the other three manufacturing domains.

Due to the fact that many of these processes are being used in scientific research labs and on non-scaleable projects, it is difficult to find consistent data. Nonetheless we found it important to understand how the nano-manufacturing field has advanced.

The most reliable metric that we found was the cost per kg of the creation of bulk carbon nano-tubes. This metric allows for the study of how much the economics of producing carbon nanotubes have improved over the last 15 years. Regrettably, this data does not include further information on the chirality, purity, diameter or length of the carbon nanotubes being produced. However, the metrics can be found for both the single wall carbon nano-tubes and the multi-wall carbon nanotubes. While this metric is incomplete, it provides interesting insight into a relatively new field that could help provide predictions for improvements in the technology for future years. Ultimately the metric for evaluating carbon nano-tubes is the cost per kg produced in inflation-adjusted USD.

#### 3.3.4.2. Carbon Nanotube Production Data

The data was collected from the text of 10 different sources, nine of which were peer reviewed journal articles. The only metric that was reported was the price per kilogram of bulk carbon nanotubes. While it would be preferred to have more metrics to describe the manufacturing of carbon nanotubes, the recency of the commercial availability of the substance makes the more limited data set acceptable. Another area where better data would be desired is in the specification of the particular types of carbon nanotubes being priced. In each of the 9 peer reviewed examples, there is no mention of purity, chirality, diameter, or length of the Carbon Nanotubes. One bright spot in the data is that we have a large number of different sources that corroborate each other and that we have data for both the production of single wall carbon nanotubes and multi-walled carbon nanotubes. The data for this domain spans from 1999 till 2013.

### 3.3.4.3. Carbon Nanotube Production technological improvement curve

Due to the limited amount of data available for the nano-manufacturing domain of carbon nano-tube production, the only metric-domain pairs are cost based. Figure 46 shows the improvement in production economics for both single and multi-walled carbon nanotubes.

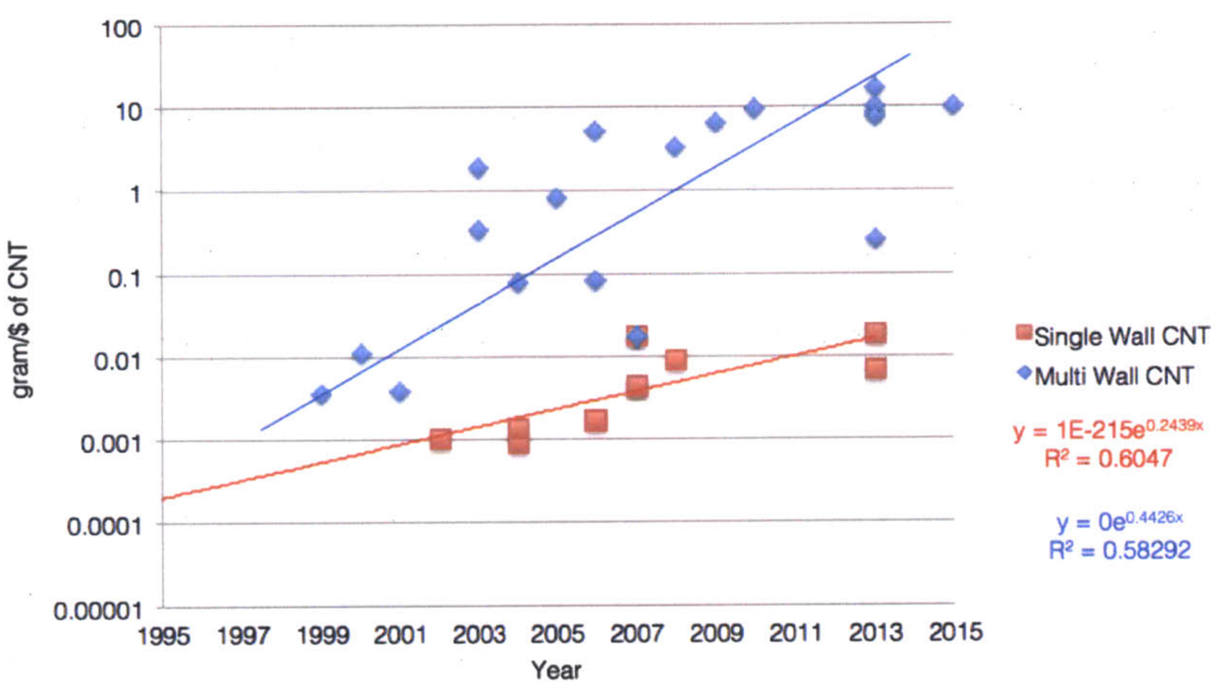



**Figure 46: Logarithm of price of single and multi-wall Carbon Nanotubes vs time with associate exponential regression lines (the blue curve is drawn in manually due to the very low coefficient)**

The improvement rate of multi-wall carbon nanotubes is 44% per year ($R^2 = 0.59$) and single wall carbon nanotubes improve at a rate of 24% per year ($R^2 = 0.6$). Multi-wall carbon nanotubes are general produced in larger quantities and are the more economical of the two choices, and appear to be improving at a more rapid pace than the more expensive single wall carbon nanotubes.

There is only measured metric for this particular domain (indicating a high likelihood of omitted variables such as length, quality, etc.), only relatively low $R^2$ values, and most importantly only a relatively short time frame, thus, there is low reliability in these improvement rates. Additionally, the nature of the data collected for these improvement curves is mostly that of laboratory uses for CNTs and not many examples of large scale commercial products. Due to the batch-like processing methods used to create CNTs – it is possible that the improvement rate will differ when the technology has reached consumer product levels – although this this pattern was not seen in the development of solar PV or other materials (Benson and Magee, 2012). Nonetheless, the wide range of sources, and the consistency of the improvement rates for the two types of carbon nanotubes indicate that CNTs are a relatively rapidly improving technological domain.

### 3.3.5. Discussion of Manufacturing TIRs

All four of the manufacturing domains showed exponential improvement of appropriate metrics. This is an important observation in that it shows consistency of the types of improvements across several different manufacturing domains. Another implication of this finding is that manufacturing domains improve similarly to those of products and processes that have been measured before, even if they are at different rates. This indicates that the development of manufacturing technological domains can be treated similarly to other types of technical domains. Had we not seen the exponential improvement, a case could have been made for exceptionalism of manufacturing domains and questions could have arisen over whether the technologies that create other domains improve in a similar fashion to the products represented in the other domains themselves.

While each of the domains exhibits exponential improvement of some metrics, there are many more domain-metric pairs that are incredibly noisy and do not show any kind of relationship between the technical metric and time. These kinds of domain-metric pairs appear more prevalently in the domains of manufacturing due to their inherent heterogeneity of outputs. It also indicates that the importance of selecting appropriate metrics is even more critical in the manufacturing domains than in other more homogenous technologies such as renewable energy generation technologies where the main output is almost always kWh. We found that in general the metrics that did follow exponential improvement rates were of the same form as the generic manufacturing metric shown in Equation 12, and that many of the metrics that were not exponential did not involve a trade-off.

Within each of the domains, there were separate rate estimates made for each of the metrics; many of the domains had very similar rates of improvement across many of the metrics. For example, all of the milling machine metrics improved at rates between 3.5% and 6.5% yearly. All of the exponential metrics for SLA 3D printers improved in the 30%-37.5% range. While there is a range for each of the domains, the general notion that SLA 3D printers improve at a more rapid rate than milling machines appears very highly likely if not certain. Although there are some differences in the improvement rates between metrics within a domain, they are much smaller than the differences in improvement rates between different domains and for milling and 3D printing the differences are probably within the high confidence accuracy range and thus cannot be distinguished. Later in the thesis, there will be a discussion regarding the domains where some metrics exhibit a faster rate of improvement than others in the same domain.

Due to the smaller ranges of intra-domain than inter-domain improvement rates, we are able to compare the relative speeds of improvement for each of the domains. The results show that milling machines improve the slowest at ~5% (with a range of +-1.5%), photolithography improves at around 24%, and is drastically reduced to only 5.3% improvement when cost is included in the metric. The fastest improvement rates were found in the production of carbon nanotubes and SLA 3D printers, which both improved at rate of ~31% (with ranges of 1% and 4% respectively). Based upon the example presented, is interesting to note that the newer manufacturing technologies improve at a rate that is higher than that of the older technologies.

## 3.4. The Classification Overlap Method

Determining appropriate TIRs for each TD provides the dependent variable for the large cross-domain experiment. We look to understand the differences in TIRs between domains through patent analysis. In order to do this effectively, it is important to locate a relevant and complete set of patents that represent a particular technological field. The relevance of a patent set resulting from a search is defined as the number of relevant patents in that set divided by the total number of patents in the same set. Similarly, completeness is the number of relevant patents in that set divided by the total number of relevant patents in the entire United States patent database (a number that can never be known for sure). This section will introduce a robust, repeatable method for selecting a set of highly relevant and complete patent sets that represent a particular TD called the Classification Overlap Method (COM). The COM is easily repeatable and can be used quickly by many different types of users, including those who are not well versed in the complexity of the patent system. This section builds off of much of the knowledge introduced in section 2.3, and describes the intuition behind the COM patent searching tool, how the method works, and the results of the COM in use for both some simple and advanced cases.

Figure 47 shows a summary of the COM method to gain more complete and relevant patent sets.

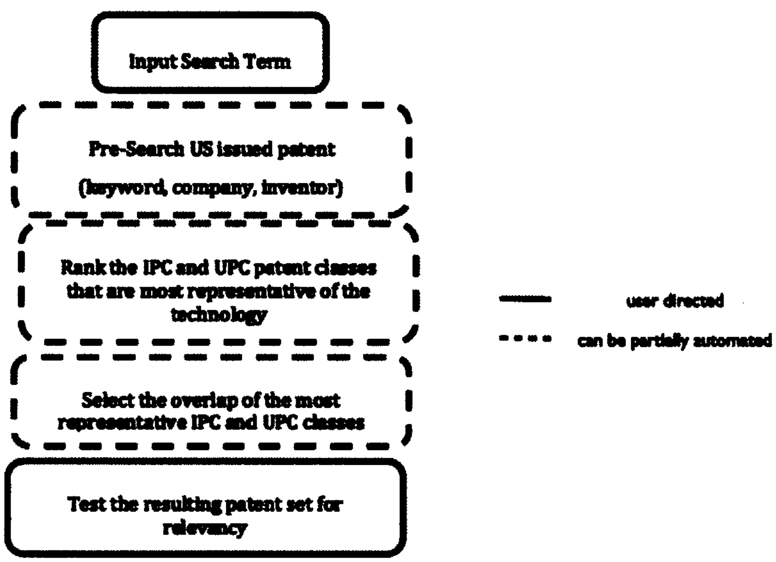



**Figure 47: Process flow of the COM: most of the method can be automated via a computer, with only the selection of the search query and the testing of the final results left to the user.**

## 3.4.1. Step 1: Pre-search US issued patent titles and abstracts for the search terms

The first step of the COM is to pre-search for the TD of interest. The most effective way to do this is to use a set of keywords to begin the process of finding the most representative patent classes (in both the US and International Patent systems), which is defined in the following section.

As one of the goals of the method is to be simple and easy to use even for someone not fluent in the patent system, the input to the COM is simply a set of search terms that can be entered into a text box. This works best with search queries of two words (ex: solar photovoltaic), which suits our use case of technological development research. The pre-search was completed by searching for the two-word query in the title or the abstract of United States Issued Patents. Thus, the pre-search identifies a set of patents with the specific query in the title or abstract.

The pre-search was done using the patent search tool PatSnap (PatSnap 2013), which searched all U. S. Patents from January 1st, 1976 to the present and was used as our database for further analysis (all of the searches in this section were completed in May of 2012 unless noted otherwise). In this paper we will give the search queries that can be used in www.patsnap.com/patents because it is publicly available and has a faster startup time than recreating a patent database from scratch. The search query used for the pre-search for 'solar photovoltaic' at www.patsnap.com/patents is:

*'TTL:(solar photovoltaic) OR ABST:(solar photovoltaic) AND DOCUMENT_TYPE:United States Issued Patent'*

This search returns 991 patents.

While the most common method of performing a pre-search for the TD is by using a keyword search, the research can be done by nearly any method. Some of the other methods that were used to locate the patent sets for this research included searching for a set of companies that are well known in the field, a set of known inventors, or even a seed set of patents. Examples of these alternative pre-search methods will be given later in the thesis when the selection of all of the patent sets is reviewed.

### 3.4.2. Step 2: Rank the IPC and UPC patent classes that are most representative of the technology

The next step in the COM method is to use the set of patents resulting from the pre-search to determine the US patent classes (UPC) and international patent classes (IPC) that are most representative of the specific technology. The representativeness ranking for the patent classes is accomplished by using the mean-precision-recall (MPR) value. This value was inspired by the 'F1' score that is common in information retrieval, but uses the arithmetic mean (instead of the geometric mean) of the precision and recall of a returned data set (Magdy and Jones, 2010). Table 6 shows an example MPR calculation for the UPCs and IPCs in the pre-search for 'solar photovoltaic'. In the paragraphs below we will describe the calculations to arrive at each column in this table and will use UPC *136* as the example.

| (1) Patent Class | (2) Number of patents in pre-search and patent class | (3) Patent Class Recall (column 2/991) | (4) Total Number of Patents in Patent Class | (5) Patent Class Precision (column 2 /column 4) | (6) MPR (column 3 + column 5)/2 |
|---|---|---|---|---|---|
| **UPC - 136** | **608** | **0.61** | **7489** | **0.081** | **0.34** |
| UPC - 257 | 342 | 0.35 | 170333 | 0.002 | 0.17 |
| UPC -438 | 178 | 0.18 | 108686 | 0.002 | 0.09 |
| UPC - 126 | 131 | 0.13 | 15450 | 0.008 | 0.07 |
| UPC -52 | 66 | 0.07 | 182645 | 0.000 | 0.03 |
| IPC - H01L | 549 | 0.55 | 180204 | 0.003 | 0.28 |
| IPC - F24J | 78 | 0.08 | 3751 | 0.021 | 0.05 |
| IPC - H02N | 47 | 0.05 | 2133 | 0.022 | 0.03 |
| IPC - H02J | 40 | 0.04 | 10876 | 0.004 | 0.02 |
| IPC - E04D | 27 | 0.03 | 3209 | 0.008 | 0.02 |

**Table 6: Example calculation of MPR for five UPCs and five IPCs for the search term 'solar photovoltaic'. The calculation for UPC 136 is described in the text and is bolded in this Table. Note that the pre-search returned 991 patents.**

Using the set of patents from the pre-search, we determine all of the unique patents classes that appear in the set. For example, within the pre-search results for 'solar photovoltaic', there are 22 unique IPCs and 10 unique UPCs. Table 6 lists the five IPCs and UPCs (column 1) with the most patents present in the search for 'solar photovoltaic'. The number of patents identified in the pre-search that are present is each class is shown in column two (this can also be called the overlap of the pre-search and patent class); it is found using a search similar to the following (using UPC 136 in this example, which returns 608 patents):

*'CCL:(136) AND TTL:(solar photovoltaic) OR ABST:(solar photovoltaic) AND DOCUMENT_TYPE:United States Issued Patent'*

Note that the sum of column two is often greater than the total number of patents in the pre-search group due to the fact that many patents are classified in multiple UPCs or IPCs.

Next, we are interested in computing the fraction of the patents in the pre-search that fall within each patent classification, also called the patent class Recall and shown in column 3 of Table 6. The recall for each of the listed patent classes is calculated by dividing the number of patents in the pre-search results that are within the patent class (column 2) by the number of patents in the pre-search patent set (991 for the example of 'solar photovoltaic'). For UPC 136, the recall is 608/991 = 0.61.

$$recall_{patentclass} = \frac{n_{pre-search\&patentclass}}{n_{pre-search}} \quad \text{(Equation 15)}$$

Next, we want to determine the total size of each of the patent classes of interest. Column number 4 shows the total number of patents in each patent class, which is found by the following search (using UPC 136 as the example, which returns 7489 patents):

*'CCL:(136) AND DOCUMENT_TYPE:United States Issued Patent'*

Given the total size of the patent class, we determine the fraction of the patents in each patent class present in the pre-search, which is called the patent class Precision (column 5). This normalizes the weight of very large and very small patent classes that may be over or under represented in the pre-search due to their different sizes. Calculate the precision of each patent class within the pre-search by dividing the number of patents in both the search and the patent class (column 2) by the total number of patents in the patent class (column 4). For UPC 136, the precision is 608/7489 = 0.081.

$$precision_{patentclass} = \frac{n_{pre-search\&patentclass}}{n_{patentclass}} \quad \text{(Equation 16)}$$

Finally, we find the mean of the precision and recall values, which gives us an estimate of how well each patent class represents the pre-search set. The MPR of each patent class (column 6) is calculated by taking the mean of the patent-class precision (column 5) and patent class recall (column 3). The MPR for UPC 136 is (0.68+0.081)/2=0.34.

$$MPR_{patentclass} = \frac{precision_{patentclass} + recall_{patentclass}}{2} \qquad \text{(Equation 17)}$$

The MPR for each potentially representative patent class - identified by containing patents present in the pre-search - are then ordered from highest MPR to lowest for both the IPC and UPC patent classification systems.

### 3.4.3. Step 3: Select the overlap of the most representative IPC class and UPC class

To find the final set, the patents that are contained within both the IPC and UPC classes with the highest MPRs within the set of U.S. issued patents are retrieved. As was introduced in section 2.3 of this thesis, our intuition for this step is founded upon the extensive US patent examiner experience and knowledge embedded in these two classification systems: the concept is to utilize that embedded knowledge. If a patent is listed in the most representative patent class in both systems (particularly since the two systems are somewhat differently structured), a reasonable hypothesis is that such dual membership results in obtaining patents of higher relevance (Criscuolo, 2006). With patents having multiple entry systems, the completeness of the set may at the same time not be too compromised. The results section will test this intuition but first we complete our description of the method.

For the solar photovoltaic case, table 7 shows the top two classes for the IPC and UPC with their corresponding MPRs as well as the size of the returned data set when the overlap of the two classes was retrieved. For example, the number of patents simultaneously contained within both highest ranked classes [136 (UPC) and H01L (IPC)] is 5101, whereas the overlap of 257 and F24J is only 16 patents, indicating quite low completeness. Selecting classes with high MPRs generally results in higher completeness and relevancy percentage combinations in the final patent set. For example, the large set (n=136406) obtained when patents that are contained within both H01L and 257 consists of a very large fraction (~98%) of irrelevant patents.

| Solar Photovoltaic' | H01L (MPR = .28) | F24J (MPR = .05) |
|---|---|---|
| 136 (MPR = .34) | 5101 | 260 |
| 257 (MPR = .17) | 136406 | 16 |

**Table 7: Comparison of Top IPC and UPC classes for the search term 'Solar Photovoltaic'; the 4 entries in the lower right hand boxes is the number of patents that are simultaneously listed in both of the specified classes (the overlap of the two classes).**

We will later discuss modifications but our direct method is to select the patents that are in both the most representative UPC and in the most representative IPC. For the solar photovoltaic example, the patent set obtained from the overlap of the most representative classes (136 and H01L) is obtained by the following query at www.patsnap.com/patents:

*'CCL:(136) OR ICL:(H01L) AND DOCUMENT_TYPE:United States Issued Patent'*

### 3.4.4. Step 4: Test the resulting patent set for relevancy

Although in this demonstration case we tested some preliminary sets for relevance, the basic process involves performing the relevance test (done by reading the abstracts of a random or semi-random test set of patents) after obtaining the set from crossing the most representative UPC class with the most representative IPC class. The relevancy sample test set size for all larger sets of patents should be 300 patents to ensure a 95% confidence interval with a margin of error of 5%. The test set structure can be varied for different purposes. For example in the case of study of technological change in a domain, we are very interested in the most highly cited patents, therefore we took the top 100 most highly cited patents and added another 200 randomly selected patents for the test set and determined the number of relevant patents using a methodology discussed in a previous paper (Benson and Magee, 2012). This relevancy testing method of the combination of the top 100 most cited patents and 200 randomly selected patents is used to evaluate the relevancy of all patent data sets in this thesis.

Depending on the use case, there may be a need for a various levels of relevance up to 100% but we consider greater than 60% of the patents being relevant acceptable for broad study of technological change. It is also important to note the absolute size of the patent sets; we found that this number varied considerably. In studying technological change over time in a domain, we are more tolerant of non-relevant patents as long as we retrieve >75% of the relevant patents (high completeness is favored over high relevance), but that may not be the case for all uses.

### 3.4.5. COM in Practice

The most important measure of effectiveness for the COM method is if it can provide highly relevant and complete patent sets for the user. Table 8 compares the overall size (an indicator of completeness) and relevancy percentage (based upon the sampling method described above) of the returned search results of three different methods across five sample technological domains. We compare the direct COM results with those resulting from a keyword search in the title and abstract because searching the total patent leads to very low relevancy and searching the title alone gives very poor completeness. Specifically for the comparative keyword searches in table 8 we use:

*'ttl:(keyword) OR abst:(keyword) AND (DOCUMENT_TYPE: United States Issued Patent)'*

This query in fact is equivalent to our "pre-search" but for the keyword search method, the resulting set of patents is the "final" set (we will examine the impact of differing search terms shortly as this has significant impact on the Keyword search patent set). Table 8 also compares the patent set that is achieved by the method of UPC classification selection. Typically, the patent classification method involves examining the UPC classification titles and making a subjective judgment on which class is best for the field of interest or possibly subjectively defining several classes of potential interest. In our comparison in table 3, we wanted a stable method so the UPC classification selected was the most representative UPC class for each field based on the objective MPR method defined above. For example, UPC class 136 – 'Batteries: thermoelectric and photoelectric' was the class identified for the 'photovoltaic electricity' case.

| Field of Interest | COM | Keyword | Classification Selection (UPC) |
|---|---|---|---|
| **Photovoltaic Electricity** | 5101 (85%) | 1006 (75%) | 7233 (57%) |
| **Wind Turbine** | 1346 (94%) | 1843 (91%) | 12893 (26%) |
| **Electric Capacitor** | 6173 (84%) | 11026 (43%) | 9472 (2%) |
| **Electrochemical Battery** | 22115 (62%) | 1159 (87%) | 26111 (62%) |
| **Computed Tomography** | 3827 (91%) | 1289 (98%) | 10444 (69%) |

**Table 8: Size and relevancy (in brackets) of returned data sets for three search methods**

The results show that the COM method does not always simultaneously produce the highest relevancy percentage and the highest estimated completeness (relevancy times size), but it consistently performs well for both characteristics and does not yield very poor results as often occurs when using the keyword or classification selection methods.

The COM search for 'Photovoltaic Electricity' provides a significant improvement over the authors' first attempt, which resulted in a set of 2484 patents with only 62% relevancy using a non-automated but more elaborate keyword search technique (Benson and Magee, 2012). In Table 8, it is noteworthy that the COM method starting from the same keyword search terms results in a patent set that is five times larger and of higher relevancy than the keyword search. The COM photovoltaic set of patents is also superior to the patent set from the third method in Table 8 as seen by relatively lower relevance for the Classification Selection set. The results of the keyword search for 'Wind Turbine' are marginally superior to those of the COM method, but the keyword search produces *much* less complete patent sets for three of the other fields of interest. Similarly, the classification selection method produces somewhat superior results for the 'Electrochemical Battery' query but remarkably poor relevance for capacitors. In the computed tomography set, the COM produces a moderately large and very relevant set of patents but does not appear to be as complete as the classification selection method.

#### 3.4.5.1. Flexibility of Search Terms and Robustness of COM Method

The COM method requires only a 2-word search term that describes the technology of interest and can take multiple synonymous or near-synonymous queries and will give the same result. For example, the search queries 'solar power' and 'photovoltaic electricity' provide the same end patent sets with COM method but very different results if one just uses a keyword search. Table 9 shows a comparison of the robustness of the COM and keyword search methods across different search terms.

| Field of Interest | HKC | Keyword | Classification Selection (UPC) |
|---|---|---|---|
| Photovoltaic Electricity | 5101 (85%) | 1006 (75%) | 7233 (57%) |
| Wind Turbine | 1346 (94%) | 1843 (91%) | 12893 (26%) |
| Electric Capacitor | 6173 (84%) | 11026 (43%) | 9472 (2%) |
| Electrochemical Battery | 22115 (62%) | 1159 (87%) | 26111 (62%) |
| Computed Tomography | 3827 (91%) | 1289 (98%) | 10444 (69%) |

**Table 9: Results of search queries for renewable energy domains and computed tomography using the COM, keyword and UPC searches**

Table 9 indicates that the COM method has a low sensitivity to the selection of initial search terms as only the term dielectric capacitor led to a substantially different set. On the other hand, the same keyword differences lead to substantial differences in the keyword search method for almost all cases. Thus, to learn more about solar photovoltaic technology, COM offers a stable (and relatively complete and relevant) patent set using a variety of different search terms whereas the keyword search data sets would be variable and of unknown quality. This lack of sensitivity to specific search terms indicates that the COM is more repeatable across different users and technical domains.

### 3.4.5.2. Advanced Uses of the COM

In arriving at the results in Table 9, the authors noted some further useful information that suggests modifications of the method for specific fields. For the patent set in the 'wind turbine' query, the 416 and 290 UPCs are almost equally representative (MPR for 416 =.45, MPR for 290 =.36). Such close comparisons occur rather often, but it many cases most of the similarly representative patent class is almost entirely present in the directly determined patent set. For example, 88% of the patents in the 204/H01M overlap are present in the 429/H01M overlap in the search for 'electrochemical cell' (MPR for 204 = .14, MPR for 429 = .37, MPR for H01M =.41). In the case of the wind energy example, there is only a 30% redundancy between the 416/F03D overlap and the 290/F03D overlap, but both patent sets are relevant to the wind energy generation field. Specifically, the 416/F03D overlap has patents that are related to the blades of a wind turbine, while the 290/F03D overlap contains patents primarily involved in the gearbox and generator portion of the wind turbines. This is also indicated by the UPC titles for the patent classes: 416 - 'Fluid reaction surfaces (i.e. impellers)' and 290 – 'Prime mover dynamo plants'. In this case, for further analysis of technological change, we recommend using both the 290/F03D overlap and the non-redundant part of the 416/F03D overlap, which, when combined, result in one patent set containing 2078 patents.

This same technique was used for obtaining a computed tomography patent set where the 378/G01N overlap, which includes 3814 patents with 76% relevancy, is combined with the original set of 378/A61B (3827/91%) to create a final data set with 7330 patents with 84% relevancy The appropriate query for the computed tomography search is:

*"CCL:(378) AND (ICL:(A61B) OR ICL:(G01N)) AND (DOCUMENT_TYPE: United States Issued Patent)"*

Another emendation suggested from the relevance test experience is further pruning of a particular overlap after the fact. This is demonstrated clearly by the search for patents related to energy storage batteries, which results in the cross of 429 – 'Chemistry: electrical current producing apparatus, product, and process' and H01M – 'Processes or means, e.g. batteries, for the direct conversion of chemical energy into electrical energy'. The unaltered set -see Table 9- results in 62% relevancy, which is only marginally adequate for our use case. During the relevancy sampling, it became clear that many of the non-relevant patents were related to fuel cells. Therefore, in order improve the patent set, we simply removed many fuel cell patents from the set by eliminating patents with fuel cell in the title -using the following query on www.patsnap.com/patents:

*"((CCL:(429) AND ICL:(H01M)) NOT (TTL:(Fuel Cell))) AND (DOCUMENT_TYPE:United States Issued Patent)"* .

The removal of the patents that had fuel cell in the title resulted in a reduction of 5649 patents, leaving a data set of 16466 patents, which, when sampled had a greatly improved 83% relevancy. This emendation helps alleviate any issues arising from very large patent classes in either the IPC or UPC with relatively small amounts of extra work by the user. Table 10 shows the comparison of the effectiveness of the different search methods including the COM method with modifications for the three fields just discussed.

| Field of Interest | COM | COM Modified | Keyword | Classification Selection |
|---|---|---|---|---|
| Photovoltaic Electricity | 5101 (85%) | 5101 (85%) | 1006 (75%) | 7233 (57%) |
| Wind Turbine | 1346 (94%) | 2078 (94%) | 1843 (91%) | 12893 (26%) |
| Electric Capacitor | 6173 (84%) | 6173 (84%) | 11026 (43%) | 9472 (2%) |
| Electrochemical Battery | 22115 (62%) | 16466 (83%) | 1157 (87%) | 26111 (62%) |
| Computed Tomography | 3827 (91%) | 7330 (84%) | 1289 (98%) | 10444 (69%) |

**Table 10: Results of the COM, the modified COM, keyword and classification searches**

When including the emendations beyond the fully automated COM method, the method becomes better than the other search methods across all of the queries we tested. While the modifications to the original COM method are helpful, they are certainly not necessary if one is interested in searching for a large number of data sets across many technical fields, as can be the case for some research related to technological development. However, the ease of making such modifications does enhance the usefulness of the COM method.

### 3.4.5.3. Comparison to an Expert Selection of Patents

While the COM method is not intended to be a replacement for an expertly selected set of patents, it is useful to understand how the method compares with a set of patents hand selected by an expert. Manuel Trajtenberg, who was widely referred to in section 2.3 of this thesis has many well cited works dealing with patents and thus can definitely be considered an expert. In order to do this, the patents from the three computed tomography searches are compared with Trajtenberg's set, using 1973-1987 as the search years in order to match the years spanning Trajtenberg's search. The results are shown in table 11 (Trajtenberg, 1990).

| | HKC | HKC Modified | Keyword | Classification Selection |
|---|---|---|---|---|
| Number of patents | 524 | 1373 | 113 | 3812 |
| Overlap with Trajtenberg | 136 | 239 | 76 | 426 |

**Table 11: Comparison of COM results to Trajtenberg's expert set (limited to years 1973-1987)**

The COM set contains 524 patents, with 136 patents overlapping Trajtentberg's set of 456 patents. If we were to assume that Trajtenberg's set is the complete set of relevant patents for the computed tomography economic domain, the COM method would have a relevancy of 26% and a completeness of 30%. The keyword search is far worse in completeness and the classification method is far inferior in relevance so the COM is the best of three weak comparators in this case. While the COM method does not match up well with the Trajtenberg set, it does manage to locate 4 of the top 5 most cited patents in the Trajtenberg set as well 3 other highly cited patents that were not in Trajtenberg's set. The results look more promising when the adjusted COM method including the 378&G01N data set is used, thereby locating 7 of the top 10 most cited patents in Trajtenberg's set.

The highly cited patents in the COM results that were not found in Trajtenberg's data set highlight the difference between searches within a technical or an economic category. Trajtenberg aimed to:

*'allow one to identify quite easily all the patents issued in predetermined economic categories, and retrieve them for further analysis.'* (Trajtenberg, 1987)

While his analysis is very similar to that of ours, we are primarily interested in identifying patents in a predetermined *technical* category as opposed to the *economic* domains that Trajtenberg focused upon. For example, the highly cited patent - number 4583242 'Apparatus for positioning a sample in a computerized axial tomographic scanner' - describes a method for locating a core sample from a borehole. This patent uses computed tomography outside of medical applications for identifying samples of rock or core for the petroleum industry. However, the patent still represents a development within the technological field. Patents such as these are clearly outside of Trajtenberg's intended field of study, but are within our broader field of study, as we are concerned with including technological spillover in our studies (Benson and Magee, 2012).

Ultimately the marginal agreement of the patents found by the COM method and the Trajtenberg set demonstrates that the COM method alone is not a replacement for an expertly selected set of patents within an economic category, but rather a robust tool to be used in conjunction with others to locate a set of patents relevant to a particular technical field.

### 3.4.6. Applying the COM to the 28 TDs

The COM provides a new tool to locate highly relevant and complete sets of patents within a technological field. The method is easily automated and straightforward to use, only requiring a query related to the field of interest. Moreover, our results show that the COM allows for flexibility in the initial keywords chosen. We have shown that the patent sets obtained from the COM method are nearly always an improvement over those obtained from the keyword or the classification search methods. Importantly, the COM method is more robust and generally easier to use. The method acts as a supplement, not a replacement for an expertly selected set of patents. COM is a simple and repeatable method for selecting sets of patents relevant to a particular technical field. The repeatability of this method should help improve consistency of patent analysis across many fields and in particular plays an important role in selecting relevant patent sets to be analyzed and compared with the TIRs of the 28 TDs.

## 3.5. Domain Patent Markers (DPMs) based on 5 Main Hypotheses

The main cross-domain experiment that this thesis is centered around involves searching for patent-based explanations for the variation of the TIR across domains. Although there is no existing theory to explain TIR differences between technological domains, there are a large number of useful theoretical writings on technological change. This section of the thesis builds upon this prior work to establish a foundation for explaining improvement rate differences as well as to develop hypotheses that are testable from the patent data. Domain Patent Markers (DPMs) are introduced that allow for the evaluation of these hypotheses using the patent sets selected by the classification overlap method for each technological domain.

### 3.5.1. Hypothesis 1: Effort

Section 2.1 of this thesis lists many theories of technological change that relate improving technology with the amount of effort that is put into a particular technology. Figure 48 shows a graphical summary of Hypothesis 1.

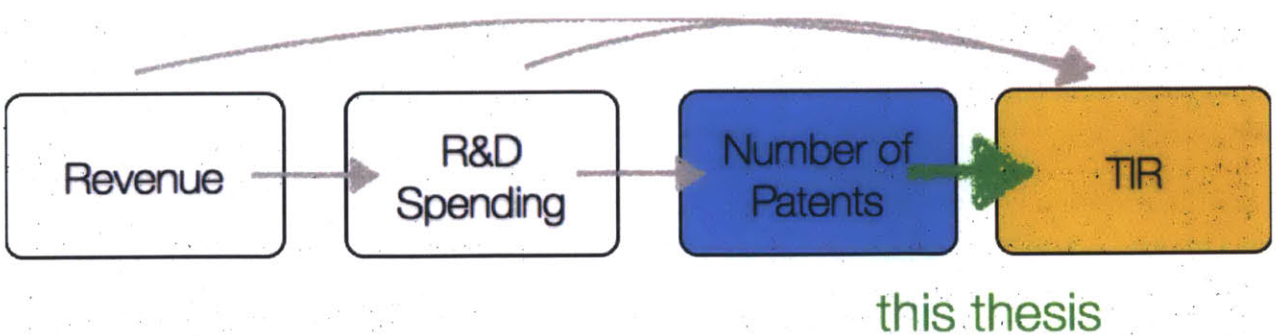



**Figure 48: Graphical Summary of Hypothesis 1 - Effort**

There are several aspects of technological evolution where the demand or usage could play an important role in the relative rate of improvement in a technological domain as are shown in figure 48 by the gray arrows and annotated letters. The theoretical basis for each of these relationships is introduced and the culmination of these relationships will be a hypothesis that can be tested using DPMs.

Annotation [A] in figure 48 represents the link between effort and technological improvement. As was mentioned previously in section 2.1.2, Wright's (1936) important paper strongly related the production (and therefore revenue) of a product with decreasing prices and contributes much of this to improvements in the related technology. This sentiment was echoed by Arrow in 1962 in his paper 'The Economic Implications of Learning by Doing' where he stated that more highly used technologies would enable more opportunity to 'learn by doing' in production as is discussed in section 2.2.4.

The relationships between the amount of revenue and the quantity of R&D spending for a technology is depicted in [B] in figure 48. As companies make more money by selling a certain product, they are likely to reinvest in that product with the hopes of making further profits from the technology. This relationship was shown by Bound et al (1982) when they confirmed a very strong correlation between the R&D efforts of a firm and their respective profits in a hundreds of firms across many industries and scales.

The direct relationship between R&D output and technical improvement is designated by [C] in figure 48 and was discussed by Christensen (1992) when he related the purely technical improvement of areal density of hard disks to the increase in engineering effort as was mentioned in section 2.2.6. The intuitive relationship between R&D effort and the number of patents produced by a particular firm is shown in [D] in figure 48 and such a relationship is supported in the work of (Margolis and Kammen, 1999). Trajtenberg's support for this idea is stated in section 2.3.

Thus, our study uses patents to test the theory that more inventive effort, as can be measured by demand, production, or R&D spending, results in better technological performance improvements. Thus the first hypothesis is:

**Hypothesis 1: The performance improvement or cost reduction rate in a technological domain should increase with the number of patents within each technological domain**

#### 3.5.1.1. DPMs to Measure Effort

Hypothesis 1 can be evaluated by simply counting the number of patents in a domain, or comparing the yearly average patenting rate for each of the domains. The DPMs used to evaluate Hypothesis 1 are listed in this section.

<u>Simple Patent Count</u>

This is the total number of patents within a technological domain. In this research, this includes patents that were published between January 1st, 1976 and July 1st, 2013. This measure is calculated using Equation 18 where SPC is the simple patent count, t is the date, and $P_t$ is the set of patents issued on that particular date, and COUNT() returns the total number of elements in a set.

$$SPC = \sum_{t=1/1/1976}^{7/1/2013} COUNT(P_t) \quad \text{(Equation 18)}$$

### 3.5.2. Hypothesis 2: Breakthroughs

Hypothesis two involves the impact of very important inventions on technological improvement. Figure 49 shows a graphical summary of the Hypothesis 2.

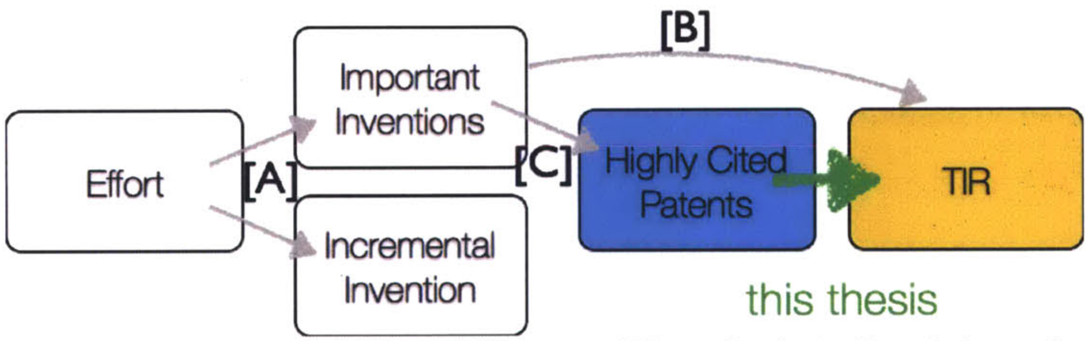



**Figure 49: Graphical Summary of Hypothesis 2 - Breakthroughs**

Section 2.2.7 discusses two type of inventions, one with a select few very important inventions and one with many inventions of lesser importance. This variation in outcome of inventive effort is depicted as [A] in figure 49. Thus, it might be reasonable to suspect that technological domains with enhanced radical invention would improve in performance faster than those with less of such breakthroughs as depicted in [B] in figure 49. This theory is supported by Tushman and Andersen (1986) who claim that the bulk of the performance improvement in a technology is driven by a few rare inventions. While it is difficult to determine objectively what qualifies as a very important invention, several studies introduced in section 2.3 show that the number of forward citations that a patent receives is likely to correlate with their importance as depicted in [C] in figure 49.

Thus, hypothesis two is as follows:

**Hypothesis 2: Technological domains with more highly cited patents should have higher rates of improvement of performance or cost reduction.**

### 3.5.2.1. DPMs to Measure Breakthroughs

The DPMs used to tests Hypothesis 2 are directly related to the future inventions that are based upon the inventions with a TD. These theories are measures of importance and impact that a field has on future inventions.

Average Number of Forward Citations per Patent

This is the average number of Forward citations for the patents in a technological domain. This measure is calculated using Equation 19 where SPC is the simple patent count, and $FC_i$ is the number of Forward citations for patent $i$. Please note that $\sum_{i=1}^{SPC}\sum_{j=1}^{FC_i} 1$ is the just the sum of the total count of Forward citations for all of the patents in the TD (without duplicates removed).

$$\frac{\sum_{i=1}^{SPC}\sum_{j=1}^{FC_i} 1}{SPC} \quad \text{(Equation 19)}$$

Total Number of Patents with more than 20 Forward Citations

This is the number of patents in a technological domain that have received more than 20 citations. This measure is calculated using equation 20 where SPC is the simple patent count, $FC_i$ is the number of Forward citations for patent *i*, and the function IF(*arg*) only counts the values if the argument is satisfied. In this situation, $IF(FC_i > 20)$ will only be counted if patent *i* has more than 20 forward citations. The specific cutoff of 20 citations is based off of the work done by Schoenmakers and Duysters (2010).

$$\sum_{i=1}^{SPC} IF(FC_i > 20) \quad \text{(Equation 21)}$$

### 3.5.3. Hypothesis 3: Science

Hypothesis 3 centers around the idea –derived from versions of the linear model discussed in section 2.2.2 that scientific progress is the basis of technological improvement and thus domains that have more reliance on basic science should improve at a faster rate than domains that are not as dependent upon science. Figure 50 shows a graphical summary of Hypothesis 3.

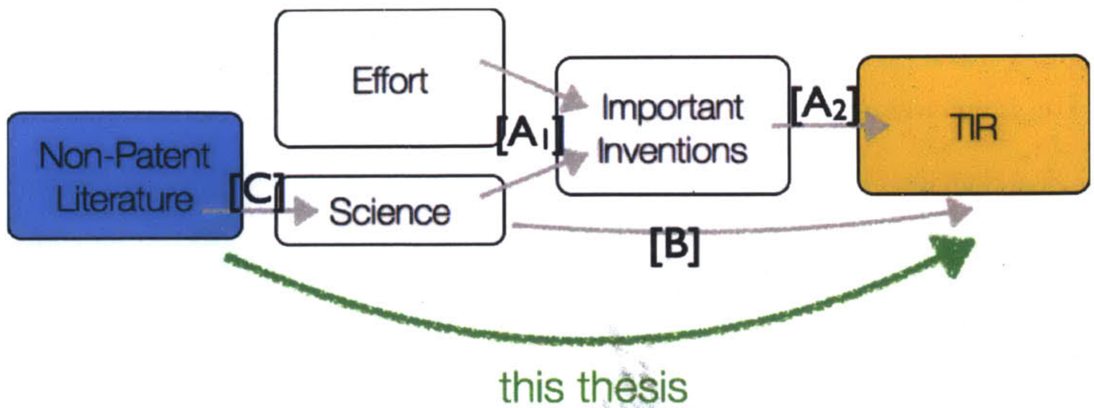



**Figure 50: Graphical Summary of Hypothesis 3 - Science**

As was discussed in section 2.3, the use of the citations in a patent that are to non-patent literature (NPL) - usually scientific journals - are available, the NPL citation fraction has been used to ascertain the scientific connection of specific patents as is shown in [C] in 50 and was demonstrated by Trajtenberg et al (1997) and others (Hall and Jaffe, 2001; Valentini, 2012)

For understanding differences in rates between domains, this theory suggests that *domains* whose patents cite more scientific articles will improve more rapidly than those who cite less such articles; the resulting hypothesis is:

**Hypothesis 3: Technological domains with a higher frequency of citations to the scientific literature should have higher rates of improvement in performance.**

### 3.5.3.1. DPMs to Measure Scientific Reliance

The DPMs used to test Hypothesis 3 are based upon the non-patent literature citations that link the patents to scientific literature.

Non-Patent Literature Citation Ratio

This is the average number of patent classes to which each patent belongs in a technological domain. This measure is calculated using Equation 22 where SPC is the simple patent count, $NPL_i$ is the number non-patent literature citations for each patent *i*, and $BC_i$ is the number of backward citations for each patent *i*.

$$\frac{\sum_{i=1}^{SPC} \frac{NPL_i}{NPL_i + BC_i}}{SPC} \qquad \text{(Equation 22)}$$

Number of Patents with >0 NPL Citation

This is the total number of patents within a domain that have at least 1 non-patent literature reference. This measure is calculated using Equation 23 where SPC is the simple patent count, $NPL_i$ is the number non-patent literature citations for each patent *i*, and IF is the operator that only count patents that have satisfied the condition - in this case of having at least 1 NPL citation.

$$\sum_{i=1}^{SPC} IF(NPL_i > 0) \qquad \text{(Equation 23)}$$

Ratio of Patents with >0 NPL Citation

This is the ratio of patents within a domain that have at least 1 non-patent literature reference. This measure is calculated using Equation 24 where SPC is the simple patent count, $NPL_i$ is the number non-patent literature citations for each patent *i*, and IF is the operator that only count patents that have satisfied the condition - in this case of having at least 1 NPL citation.

$$\frac{\sum_{i=1}^{SPC} IF(NPL_i > 0)}{SPC} \qquad \text{(Equation 24)}$$

### 3.5.4. Hypothesis 4: Recency

The ideas that more rapidly improving domains are newer and are based upon more recent knowledge (and the resulting feedback loop) forms the basic intuition for Hypothesis 4. More specifically, we examine whether domains that are newer or rely upon more recent knowledge improve at a more rapid pace than their older counterparts. Figure 51 shows a graphical summary of Hypothesis 4.

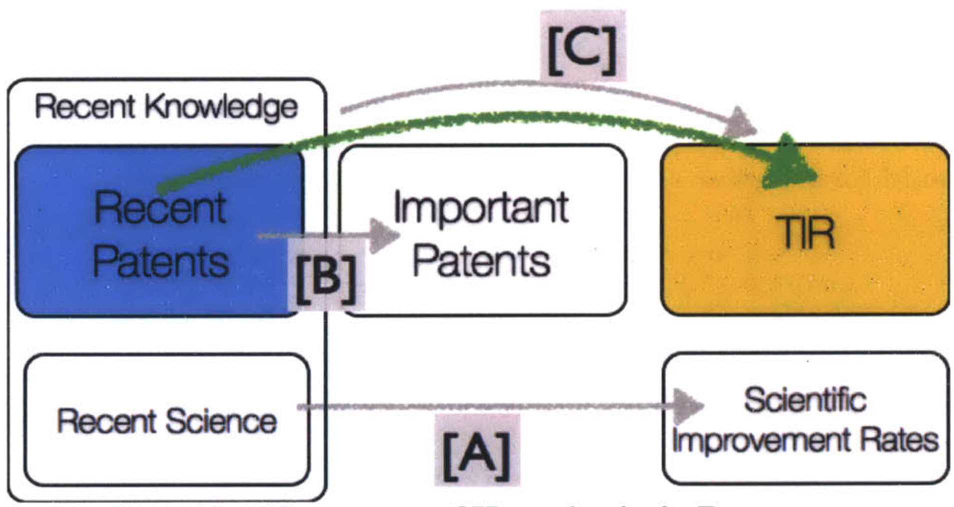



**Figure 51: Graphical Summary of Hypothesis 4 - Recency**

While not specifically related to technological improvement, the relationship between more recent science and more rapidly improving scientific fields as depicted in [A] in figure 51 provides a promising analogy for the recency of patents and technological improvement. The connection between recency of science and higher scientific improvement rates was studied in depth by Price (1965), who showed that fast improving scientific fields follow a 'research front' from that relies mainly on very recently published papers. Schoenmakers and Duysters (2010) showed that more important inventions tended to rely upon newer technologies as depicted in [B] in figure 51. This relationship would indicate that recent inventions act a root cause of technological improvement. Another explanation of these relationships is that recent knowledge could act as an instrumental variable and cause both technological improvement and the creation of important inventions as depicted in [C] in figure 51. Such a relationship was shown by Nerkar (2003) in his study of path-dependency in inventions.

The logic then follows that if more recent patents are the most important ones to current patents, acceleration of performance is more likely than for a field that is still relying on older foundations; the resulting hypothesis is:

**Hypothesis 4: Technological domains whose patents cite more recent patents should have higher rates of improvement in performance.**

### 3.5.4.1. Recency DPMs

Hypothesis 4 is evaluated using DPMs that are directly related to the dates of publication of the TD patent sets and its related citations. These metrics provide an overview of age-related characteristics for each TD.

Average Publication Year

This is the average year of publication for the patents within a technological domain. In this research, this includes patents that were published between January 1st, 1976 and July 1st,

2013. This measure is calculated using Equation 25 where SPC is the simple patent count and $t_{i_{pub}}$ is the publication year of patent *i*.

$$\frac{\sum_{i=1}^{SPC} t_{i_{pub}}}{SPC} \qquad \text{(Equation 25)}$$

Price Index (3 years)

This is the average number of forward citations that each patent received within 3 years of publication for patents in a technological domain divided by the average total number of forward citations per patent. This measure is calculated using Equation 26 where SPC is the simple patent count, $FC_i$ is the number of Forward citations for patent *i*, $t_{i_{pub}}$ is the publication year of patent *i*, $t_{ij_{pub}}$ is the publication date of forward citation *j* of patent *i*, and the function IF(*arg*) only counts the values if the argument is satisfied.

$$\frac{\sum_{i=1}^{SPC}\sum_{j=1}^{FC_i} IF(t_{ij_{pub}} - t_{i_{pub}} \le 3)}{\frac{\sum_{i=1}^{SPC}\sum_{j=1}^{FC_i} 1}{SPC}} \qquad \text{(Equation 26)}$$

The next two DPMs *are a combination of hypothesis 2 and hypothesis 4, as they are measures of how many citations a patent receives soon after publication.*

Average number of Forward Citations within 3 years of publication per patent

This DPM is simply the numerator of the price index (3 years) listed above. This is the average number of forward citations that each patent received within 3 years of publication for patents in a technological domain. This measure is calculated using Equation 27 where SPC is the simple patent count, $FC_i$ is the number of Forward citations for patent *i*, $t_{i_{pub}}$ is the publication year of patent *i*, $t_{ij_{pub}}$ is the publication date of forward citation *j* of patent *i*, and the function IF(*arg*) only counts the values if the argument is satisfied.

$$\sum_{i=1}^{SPC}\sum_{j=1}^{FC_i} IF(t_{ij_{pub}} - t_{i_{pub}} \leq 3) \qquad \text{(Equation 27)}$$

Average Age of Backward Citations

This is the average age of publication of the backward citations for patents within a technological domain. This measure is calculated using Equation 28 where SPC is the simple patent count, $BC_i$ is the number of backward citations for patent *i*, $t_{ji_{pub}}$ is the year of publication of backward citation *j* of patent *i* and $t_{i_{pub}}$ is the publication year of patent *i*. Note that this equation is the average publication date - the average publication date of backward citations.

$$\frac{\sum_{i=1}^{SPC} t_{i_{pub}}}{SPC} - \frac{\sum_{i=1}^{SPC}\sum_{j=1}^{BC_i} t_{ji_{pub}}}{\sum_{i=1}^{SPC}\sum_{j=1}^{BC_i} 1} \qquad \text{(Equation 28)}$$

### 3.5.5. Hypothesis 5: Breadth of Knowledge

The intuition behind hypothesis 5 is that domains that rely upon knowledge from a broader knowledge base are likely to improve more quickly and are discussed in depth in section 2.2 of this thesis. Figure 52 shows a graphical summary of the Hypothesis 5.

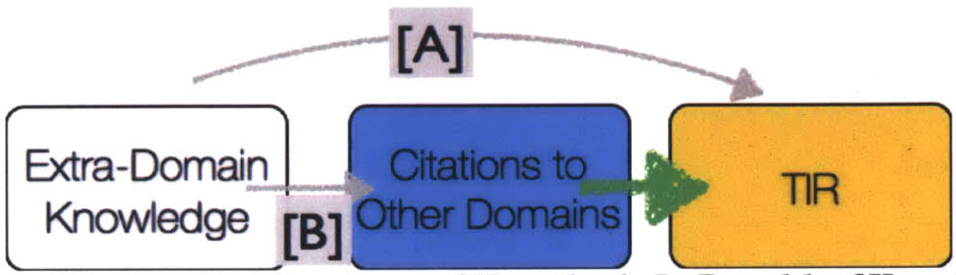



**Figure 52: Graphical Summary of Hypothesis 5 - Breadth of Knowledge**

The theories regarding breadth of knowledge are often related to the combining of knowledge from different domains, stating that the use of information from very different sources is likely to result in improved technological outcomes as is shown by [A] in figure 52. Rosenberg (1982) showed that technological spillover greatly impacted the quantity and quality of technological change in the United States in the 20th century. A recent paper by Nemet and Johnson (2012) reviews the studies that link the flow of knowledge between inventions to citation flows as shown by [B] in figure 52.

*Many studies have found that citations from one patent to another provide a means for measuring the flows of knowledge (Trajtenberg, 1990; Jaffe et al., 1998, 2000). Detailed technology case studies of citation networks generally find that cited prior does include precursor inventions (Mina et al., 2007; Fontana et al., 2009; Barbera-Tomas et al., 2011). (Nemet and Johnson, 2012)*

This link between citations and flows of knowledge is crucial to several of the domain patent markers that are used to test these hypotheses. Nemet and Johnson(2012) find that their results do NOT support the theory that important inventions involve the transfer of new knowledge from one technological domain to the other through patents. Thus hypothesis 5 is:

**Hypothesis 5: Technological domains that cite higher internal fractions of patents from their own domain will have higher rates of improvement.**

### 3.5.5.1. DPMs to Measure Breadth of Knowledge

The DPMs that are used to evaluate Hypothesis 5 are related to the knowledge base on which a TD is built. These theories are intended to find the roots of the information that was used to create the inventions in the TD.

Percentage of Backward Citations to Own Domain

This is the percentage of backward citations from the entire patent set that are to patents within the patent set. This measure is calculated using Equation 29 where SPC is the simple patent count, and $BC_i$ is the set of backward citations for patent *i*, $P_i$ is the total set of patents within the TD and $\bigcup$ is the union of two sets, $\bigcap$ is the intersection between two sets and COUNT() counts the number of elements in a set.

$$\frac{COUNT(\bigcup_{i=1}^{SPC} P_i \bigcap \bigcup_{i=1}^{SPC} BC_i)}{SPC} \quad \text{(Equation 29)}$$

### 3.5.6. Other Domain Patent Markers

While most of the DPMs in this study were developed specifically to test the 5 main hypotheses, several other DPMs were used as additional tests for correlation with TIRs of the TDs. This subsection contains DPMs that are directly related to the biographical data of a patent and the TD patent set. Some of the DPMs in this category can act as controls as they are not supported by any particular theory. Some additional technological change theories that can be evaluated using these metrics are as follows:

- A wider variety of human input for a technology (more inventors or assignees) can lead to better inventions and a higher TIR
- Inventions that have multiple forms (many family members) are likely to lead to more innovation and thus a higher TIR
- Patents with longer descriptors (title, abstract claim) are more detailed and thus are more focused, and this focus leads to a higher TIR
- Conversely patents with shorter descriptions are more general and thus enable more inventions and a higher TIR

Average Number of Inventors per Patent

This is the average number of inventors that are listed on each patent within a technological domain. This measure is calculated using Equation 30 where SPC is the simple patent count and $inventor_i$ is the count of the number of inventors for each patent *i*.

$$\frac{\sum_{i=1}^{SPC} inventor_i}{SPC} \quad \text{(Equation 30)}$$

Average Number of Assignees per Patent

This is the average number of assignees that are listed on each patent within a technological domain. This measure is calculated using Equation 31 where SPC is the simple patent count and $assignee_i$ is the count of the number of assignees for each patent *i*.

$$\frac{\sum_{i=1}^{SPC} assignee_i}{SPC} \quad \text{(Equation 31)}$$

Average Number of INPADOC Family Members Per Patent

This is the average number of INPADOC family members that are listed on each patent within a technological domain. INPADOC refers to the database that is being used to populate the information and has a definition of patent 'family members' that refers to documents that are all linked through a common priority filing document.[3] This measure is calculated using Equation 32 where SPC is the simple patent count and $INPADOC_i$ is the count of the number INPADOC family members for each patent *i*.

$$\frac{\sum_{i=1}^{SPC} INPADOC_i}{SPC} \quad \text{(Equation 32)}$$

The next three DPMs are preliminary attempts at textual analysis. Full-scale semantic or textual analysis is beyond the scope of this thesis, but is certainly an opportunity for future research that is discussed in the end of the thesis.

Average Length (in characters) of the Title per Patent

This is the average number of characters (including spaces) of the title of each patent within a technological domain. This measure is calculated using Equation 33 where SPC is the simple patent count and $Length(title)_i$ is the count of the number of characters in the title for each patent *i*.

$$\frac{\sum_{i=1}^{SPC} Length(title)_i}{SPC} \quad \text{(Equation 33)}$$

Average Length (in characters) of the Abstract per Patent

This is the average number of characters (including spaces) of the abstract of each patent within a technological domain. This measure is calculated using Equation 34 where SPC is the simple patent count and $Length(abstract)_i$ is the count of the number of characters in the abstract for each patent *i*.

$$\frac{\sum_{i=1}^{SPC} Length(abstract)_i}{SPC} \quad \text{(Equation 34)}$$

Average Length (in characters) of the First Claim of each Patent

This is the average number of characters (including spaces) of the first claim of each patent within a technological domain. This measure is calculated using Equation 35 where SPC is the simple patent count and $\mathit{Length(abstract)}_i$ is the count of the number of characters in the first claim for each patent *i*.

$$\frac{\sum_{i=1}^{SPC} \mathit{Length(abstract)}_i}{SPC} \qquad \text{(Equation 35)}$$

## 3.6. Renewable Energy Case Study

This section summarizes a recent publication (Benson and Magee, 2014) based upon this thesis that presents four renewable energy technologies as a short demonstration to show how the TIRs can be related to the DPMs. Only one DPM from each hypothesis is be used to demonstrate the overall concept.

Two pairs of renewable energy technologies were analyzed. Importantly, the annual improvement rate of cost/investment is quite different for the four technological domains: namely, solar photovoltaics (PV) (9.5% per year), wind turbines (9.2%), batteries (7.0%) and capacitors (14.6%). While examining these TIRs for individual domains is important, this paper addresses the relative rate of cost reduction in groups of competing technologies. Among competitive approaches, those improving faster than the alternatives that are available are likely to be most economically viable and thus most highly used in the longer term. There are significant differences in rates between the selected sets of technologies, and if indicative of the future could determine which domains end up dominating their respective function (energy generation, energy storage). For the chief purpose of this section the large differences between the improvement rates sufficiently differentiates the four domains to empirically examine the hypotheses that were discussed in section 3.5 of this thesis. Table 12 shows the summary of the results for the applicable measure for each hypothesis.

| Domain | **Solar PV** | **Wind** | **Capacitors** | **Batteries** |
|---|---|---|---|---|
| Improvement Rate | 9.5% | 9.2% | 14.0% | 7.0% |
| H1 - Simple Patent Count | 5203 | 2498 | 5944 | 16122 |
| H2 - % of patents cited over 20 times | 26% | 19% | 17.5% | 15.6% |
| H3 - NPL Citation % | 22% | 10% | 11% | 18% |
| H4 - Average age of cited patents (years) | 10.6 | 17.3 | 10.2 | 10.45 |
| H5 - % of cites to own domains | 8.3% | 6.3% | 9.4% | 10.1% |

**Table 12: Short case study of TIR-DPM comparison for 4 Renewable Energy Domains**

It is informative to compare the results by looking at the four domains as two independent sets (one for energy generation and one for energy storage). There are only two hypotheses that are consistent with the empirical results for both pairs in this type of comparison. The proportion of highly cited patents is higher for the two technologies with higher improvement rates (solar PV and capacitors) in accord with Hypothesis 2. This may lend credence to the idea that radical or breakthrough inventions tend to move technologies forward, although the difference between batteries and capacitors is very small for this metric, which weakens the argument since the battery domain progresses much more slowly than the capacitor domain. Another signal in this comparison is that from the average ages of cited patents which is smaller for solar and capacitors as suggested by hypothesis 4. These results support the idea that domains that rely on more recent technology tend to develop more quickly.

In order to analyze the four data points as a whole, we performed a correlation analysis on this data. Not surprisingly, the limited number of data points in the publication case study (four) assures that there were no statistically significant correlations. In fact, the qualitative review agrees with the statistical test in showing no reliable explanation for the difference in rates; this review provides very weak support for Hypothesis 2 - Breakthroughs and Hypothesis 4 - Recency. The lack of statistical significance in such a small sample requires the analysis of a large number of domains, hence the broad application to 28 TDs reported in this thesis.

## 3.7. Statistical Comparison of TIRs and DPMs

The structure of the large cross-domain experiment attempts to explain the variation in TIRs (the dependent variable) with the variation in the various DPMs (independent variables). The objective is to determine which of the DPMs correlate significantly with the TIRs. A DPM that correlates positively and significantly with the TIR is support for the related hypothesis.

Due to the large number of independent variables (DPMs) it is possible that even a statistically significant correlation could be a false positive. In an attempt to ameliorate this risk, several robustness tests were performed. Most of the hypotheses have more than one DPM that is used to evaluate the hypothesis, thus a single positive (or negative) indication will not be the sole determinant of support (or lack thereof) of the hypothesis. Additionally, domain-based and time-based robustness tests have been used to validate the results and reduce the likelihood of results that are based purely on the particular domains or the associated time frame that are used in the test.

The DPMs that are most strongly correlated with TIR and prove robust to time and domain-sensitivity are then combined into predictive models of TIR using linear regression. The statistical tests associated with these models provide evidence to support or deny their predictive applicability. In addition to the primary regressions that contain the strongest DPMs, a number of additional regression models were tested using the other DPMs to ensure that the strongest DPM-based predictive model was located.

# Chapter 4: Results

This section of the thesis gives the results of the cross-domain comparison of the domain patent markers and the technological improvement rates. Before giving these key findings, the results for technological improvement rates and patents that are essential to the cross-domain comparison are shown. Following the methodology outlined in Chapter 3, this section will begin with a discussion of the technological improvement rate for each of the domain-metric-pairs, including the selection of relevant functional performance metric and the statistical robustness tests for the technological improvement rates. Next, the patent sets for each of the TDs that were selected using the COM will be discussed, along with the initial pre-searches and some of the emendations of the COM that were used to find highly relevant and precise patents sets. Next, the results of applying the domain patent marker to each of the TDs will be shown, including the range of each. Finally, the domain patent markers will be related to the technological improvement rates through a set of correlation tables and statistical regressions, including robustness tests of random sets of TDs and time sensitivity analysis to understand the potential predictive capability of the measures. This Chapter will intentionally avoid deep interpretation of the results, as that topic will be covered in Chapter 5 of the thesis.

## 4.1. Overview of Technological Improvement Rates

In total, 28 TDs resulted in 88 domain metric pairs when accounting for all of the different potential metrics for each technology. Data was collected for each of the domain metric pairs and a technological improvement rate was then derived from these data sets as was described in section 3.2.1. This subsection will describe each of the TDs along with their corresponding domain metric pairs and technological improvement rates. Finally, the results of statistical tests introduced in section 3.2.4 for each of the technological improvement rates will be discussed along with the selection of the final technological improvement rates for each of the 28 TDs that are compared with the domain patent markers to determine the final correlation of the rates.

### 4.1.1. Domains and Functional Performance Metrics (FPMs)

The 28 TDs were analyzed and a resulting in a total of 88 domain metric pairs were derived from FPMs that should represent an aspect of the consumers purchasing decision. The following sub-section will explain each of the domains along with their resulting FPMs. This sub-section should read slightly like a list of all of the domains along with their relevant FPMs. Each of the FPMs listed in this section will correspond to a domain metric pair and technological improvement rate in the next sub-section. The units of each FPM along with their derivation are also listed and although many are variations on SI units, some of the FPMs are kept in units that are common to that particular industry.

The TDs are listed in alphabetical order and each of the FPMs is listed in approximate order of increasing completeness for each TD. Additional explanation is given for FPMs that are particularly uncommon or very domain-specific (e. g., Computed Tomography). While it is ideal to find the most complete FPM for each domain, often times the very complete FPMs are limited by data availability and thus the combination of the less complete FPMs resulted in an unreliable TIR. Finally, when necessary, potentially important omitted variables for a technological domain will be mentioned after the description of the FPMs.

### 3D-Printing (industrial stereolithography)

$$\frac{speed(\frac{mm}{sec})}{layerthickness(mm)} = \frac{1}{sec}$$

- The simplest measure of SLA 3D printing is how fast the laser moves when curing the resin while maintaining a specific layer thickness. It must be noted that in some cases, such as this one the individual metrics can be coupled. For example, with fixed laser power, one would maximize layer thickness by increasing speed and decreasing layer thickness (because the laser would spend less time curing each unit of resin as it moves faster).

$$\frac{speed(\frac{mm}{sec})}{layerthickness(mm)*cost(\$)} = \frac{1}{sec*\$}$$

- Another measure of SLA 3D printing is how fast the laser moves when curing the resin while maintaining a specific layer thickness at a certain cost to the consumer.

$$\frac{speed(\frac{mm}{sec})*buildvolume(mm^3)}{layerthickness(mm)*machinesize(mm^3)*cost(\$)} = \frac{1}{sec*\$}$$

- The more complete measure of SLA 3D printing takes into account the build volume size, which allows for the production of larger parts and also the size of the machine, for which space can be at a premium in labs or production centers.

### Aircraft Transport

$$\frac{passenger*miles(mi)}{hour(hr)} = passenger*mph$$

- Aircraft transport has a commonly used metric that takes into account the speed of an aircraft along with the number of passengers an aircraft is capable of transporting. This is basically a measure of the productivity of an aircraft for a purchasing airline (Martino, 1971).

Camera Sensitivity

$\frac{saturation\_output(mV)}{pixelsize(\mu m^2)} = \frac{mV}{\mu m^2}$ - The measurement for camera sensitivity involves the saturation output measured in mV per square area of a pixel.

Combustion Engines

$\frac{power(W)}{cost(\$)} = \frac{W}{\$}$ - A measure of the performance of a combustion engine is the amount of power it produces per cost of the engine.

$\frac{power(W)}{volume(L)} = \frac{W}{L}$ - Another measure of the performance of a combustion engine is the amount of power it produces divided by the overall volume of the engine.

$\frac{power(W)}{mass(kg)} = \frac{W}{kg}$ - Another measure of the performance of a combustion engine is the amount of power it produces divided by the overall mass of the engine.

Within the combustion engine TD, the data collected could be split into several sub-domains. TIRs were collected for each of the following sub-TDs:

Aircraft Piston combustion engine

Aircraft Turbine combustion engine

Automobile Piston combustion engine

The final TIRs that were used in the comparison with the domain patent marker were a combination of all of the metrics.

Computed Tomography (CT)

$$\frac{distinct\_detail}{image\_depth(mm) * scan\_time(\sec)} = \frac{1}{mm * \sec}$$

The metrics used in CT scanning are slightly more domain-specific than most, and Peterschmitt (2007) gives a good summary of how a CT scanner is measured.

*The main purpose of medical imaging is to provide access to a reliable diagnosis method to a large set of people. An efficient medical imaging device can ideally help diagnose quickly and accurately cancers (for instance) to attempt to have an impact on care quality:*

*- The image provided by the imaging device has to show sufficiently small and distinct details. Intuitively, the spatial resolution of an imaging system can be defined in terms of the smallest spacing between two objects that can be imaged clearly. Resolution is therefore an adapted criterion. The number of distinct detail per unit length can alternatively be chosen to show an improvement in technology.*

*- The studied technique should be convenient for the patient: a short examination free of undesirable side effects is a second quality criterion. Evaluating such a characteristic is not obvious and many newly developed medical imaging techniques brought about long exposure time and induced important radiation absorbed dose for patients and radiologists. Indeed, those long exposure to ionizing radiation beams cause tissue damage. Hence, the time needed to produce a single image is also an adapted criterion. Inversely, the frequency with which images are*

*processed matters. The two criteria can be linked in a single FPM: the number of distinct detail per millimeter per second. (Peterschmitt, 2007)*

## Capacitor Energy Storage

$$\frac{energy(kWhr)}{cost(\$)} = \frac{kWhr}{\$}$$ - A measure of the performance of a capacitor used for energy storage is the amount of energy it can store per unit cost.

$$\frac{energy(kWhr)}{volume(L)} = \frac{kWhr}{L}$$ - Another measure of the performance of a capacitor used for energy storage is the amount of energy it can store per unit volume.

$$\frac{energy(kWhr)}{mass(kg)} = \frac{kWhr}{kg}$$ - Another measure of the performance of a capacitor is the amount of power it produces divided by the overall mass.

Other FPMS could include the combination of all of these variables, resulting in a variable such as $\frac{energy(kWhr)}{volume(L) * mass(kg) * cost(\$)}$, however data availability made the TIR for that very complete FPM unreliable and there is additional issues with variables that are near equals such as volume and mass (which should improve together with constant density). Potential omitted variables that may also influence the purchasing decision of a capacitor are the ability to charge and discharge rapidly, which can be measured by power (W).

## Electric Motors

$\frac{power(W)}{volume(L)} = \frac{W}{L}$ - A measure of the performance of an electric motor is the amount of power it produces divided by the overall volume.

$\frac{power(W)}{mass(kg)} = \frac{W}{kg}$ - Another measure of the performance of an electric motor is the amount of power it produces divided by the overall mass.

Electrical Energy Transmission

$power(W) * distance(km) = W * km$ - Electrical energy transmission can be measured by the amount of power transmitted over a distance.

$\frac{power(W) * distance(km)}{cost(\$)} = \frac{W * km}{\$}$ - Electrical energy transmission can be measured by the amount of power transmitted over a certain distance divided by the cost. A consumer of electrical energy transmission technologies will want to transmit a large amount of power over a long distance at a small cost.

These same FPMs can be used to measure AC and DC energy transmission, and the FPM used in the TIR/DPM comparison uses the union of the two data sets.

Electrical Information Transmission

$\frac{information(kB)}{time(\text{sec})} = \frac{kB}{\text{sec}}$ - Transmitting information over electrical (coaxial) cables can be measured by the amount of information transmitted per unit time.

$\frac{information(kB)}{time(\sec)*cost(\$)}=\frac{kB}{\sec*\$}$ - Another measure for the transmission of information over electrical (coaxial) cables is the amount of information transmitted per unit time per unit cost.

Electrochemical Battery Energy Storage

$\frac{energy(Whr)}{cost(\$)}=\frac{Whr}{\$}$ - A measure of the performance of an electrochemical battery used for energy storage is the amount of energy it can store per unit cost.

$\frac{energy(Whr)}{volume(L)}=\frac{Whr}{L}$ - Another measure of the performance of an electrochemical battery used for energy storage is the amount of energy it can store per unit volume.

$\frac{energy(Whr)}{mass(kg)}=\frac{Whr}{kg}$ - Another measure of the performance of an electrochemical battery is the amount of power it produces divided by the overall mass.

Electronic Computation

$\frac{computations}{time(\sec)}=\frac{1}{\sec}$ - Electronic computations can be measured by their speed, which is calculated by the number of computations per second.

$\frac{computations}{time(\sec)*cost(\$)}=\frac{1}{\sec*\$}$ - Another FPM includes cost in the calculation for a more complete metric.

Potential omitted variables include energy used, this metric has become especially important in recent years with the increased prevalence of mobile computing.

Flywheel Energy Storage

$$\frac{energy(kWhr)}{mass(kg)} = \frac{kWhr}{kg}$$

- Using the same metric as batteries and capacitors, energy storage can be measured by the amount of energy stored by the total mass of the flywheel. Sufficient data was not available for the other two FPMs used for batteries and capacitors (cost and volume) and therefore were not included in the TIRs for flywheel energy storage.

Fuel Cell Energy Production

$$\frac{peak\_power(kW_{peak})}{cost(\$)} = \frac{kW_{peak}}{\$}$$

- Fuel cells can be measured by the maximum amount of power produced divided by the cost of the fuel cell system.

Potential omitted variables include measures of system size include mass (kg) and volume (L).

Genome Sequencing

$$\frac{basepairs}{cost(\$)} = \frac{1}{\$}$$

- Sequencing of genomes can be measured by the number of base-pairs that can be decoded per $. A potential omitted variable that may be included is also the time in which it takes to decode the base pairs, which at this point has very few available data points and was not included in this thesis.

Incandescent Artificial Illumination

$$\frac{1000 * lumens(lm) * time(hr)}{cost(\$)} = \frac{1000 * lm * hr}{\$}$$

- A measure of illumination for incandescent light bulbs is the brightness of a bulb measured in lumens by the length of time it lasts (hrs) divided by the cost, in this case, the FPMs is measured in thousands of lumen-hours per dollar.

Integrated Circuit Information Storage

$$\frac{transistors}{die} = \varnothing$$

- The measurement of integrated circuit technologies ultimately evolves from Moore's law and thus the measurement of integrated circuit memory chips uses the same metric of the number of transistors per die.

Integrated Circuit Processors

$$\frac{transistors}{die} = \varnothing$$

- The measurement of integrated circuit technologies ultimately evolves from Moore's law and thus the measurement of integrated circuit processors uses the same metric of the number of transistors per die.

LED Artificial Illumination

$$\frac{lumens(lm)}{lamp} = lm$$

- LED lights can be measured by the amount of light that is produced by one lamp/bulb.

$$\frac{lumens(lm)}{cost(\$)} = \frac{lm}{\$}$$

- LED lights can be measured by the amount of light that they produce by the cost of the light.

Magnetic Resonance Imaging (MRI)

$$\frac{distinct\_detail}{image\_depth(mm) * scan\_time(\text{sec})} = \frac{1}{mm * \text{sec}}$$

- The metrics used in MRI machines are similar to that of CT scanners and are a measure of detail of the scan along with the scan time. See the CT scanner FPM description for a more detailed description of the metric give by Peterschmitt (2007).

$$\frac{distinct\_detail}{image\_depth(mm) * scan\_time(\text{sec}) * cost(\$)} = \frac{1}{mm * \text{sec} * \$}$$

- The metrics used in MRI machines are similar to that of CT scanners with the inclusion of cost in the denominator. See the CT scanner FPM description for a more detailed description of the metric give by Peterschmitt (2007).

Magnetic Information Storage

$$\frac{megabits(Mb)}{cost(\$)} = \frac{Mb}{\$}$$

- Magnetic information storage can be measured by the amount of information stored (measured in bits) per unit cost. Please note that 1 bit = 8 bytes.

$$\frac{mbits(Mb)}{volume(cm^3)} = \frac{Mb}{cm^3}$$

- Magnetic information storage is measured by the amount of information stored (measured in bits) per unit volume ($cm^3$). Please note that 1 bit = 8 bytes.

As in some of the other TD, magnetic information storage can be broken down into smaller sub-domains. Magnetic information has been stored on magnetic tape and magnetic hard disks, the TIR used in the TIR/DPM comparison is the combination of both metrics.

Milling Machines

$\frac{power(hp)}{accuracy(mm)} = \frac{hp}{mm}$ - Milling machine capability can be measured by the power of a machine at a certain accuracy. The power of the machine is in this case a measure for speed as was explained in more depth in the methodology section case study of manufacturing technologies. It is important to note that this is the domain that used the less reliable yearly average and the DAYC methods in constructing the TIRs.

Optical Information Storage

$\frac{megabits(Mb)}{cost(\$)} = \frac{Mb}{\$}$ - Optical information storage can be measured by the amount of information stored (measured in bits) per unit cost. Please note that 1 bit = 8 bytes.

$\frac{mbits(Mb)}{volume(cm^3)} = \frac{Mb}{cm^3}$ - Optical information storage is measured by the amount of information stored (measured in bits) per unit volume ($cm^3$). Please note that 1 bit = 8 bytes.

Optical Information Transmission

$\frac{kilobits(Kbits)}{time(\sec)} = \frac{Kbits}{\sec}$ - Optical information transmission can be measured by the bandwidth, which is measured in the amount of information transmitted per unit time. Please note that 1 bit = 8 bytes.

$\frac{kilobits(Kbits)}{time(\sec) * cable\_length(km) * cost(\$)} = \frac{Kbits}{\sec * km * \$}$ - A more thorough measure is one that calculates the bandwidth per distance of cable and at a certain cost. Please note that 1 bit = 8 bytes.

## Photolithography

$$\frac{1}{accuracy(\mu m) * cost(\$)} = \frac{1}{\mu m * \$}$$

- A measure of performance for photolithography is the accuracy of the process (1/nm) per unit cost. The measure of accuracy is described in detail in section 3.4 of the thesis. Of particular note is that the cost is the tool cost and does not take into account other costs such as the mask and energy.

$$\frac{areal\_throughput(\frac{in^2}{hr})}{cost(\$)} = \frac{in^2}{hr * \mu m * \$}$$

- Another measure of photolithography takes into account the speed at which the process operates called areal throughput divided by the cost. The areal throughput parameter is a modification of the industry-accepted metric of wafers/hr multiplies by the area of the wafer.

$$\frac{areal\_throughput(\frac{in^2}{hr})}{accuracy(\mu m)} = \frac{in^2}{hr * \mu m}$$

- Another measure of photolithography takes into account the speed at which the process operates called areal throughput divided by the accuracy. The areal throughput parameter is a modification of the industry-accepted metric of wafers/hr multiplies by the area of the wafer.

$$\frac{areal\_throughput(\frac{in^2}{hr})}{accuracy(\mu m) * cost(\$)} = \frac{in^2}{hr * \mu m * \$}$$

- The most complete measure for photolithography takes into account the speed of producing wafers, the accuracy at which they are created and the cost of doing so.

## Solar Photovoltaic Energy Generation

$$\frac{average\_energy\_produced(kWhr)}{cost(\$)} = \frac{kWhr}{\$}$$

- A measure of the average energy output of solar PV panels per unit cost also takes into account such factors as placement, tracking of the sun, and other factors that are often combined into one measure called the 'fill factor'.

$$\frac{peak\_power(W_{peak})}{cost(\$)} = \frac{W_{peak}}{\$}$$

- A measure of the device performance of a solar PV module is the peak power per unit cost.

## Superconductivity

$temperature_{critical}(K) = K$ - The measure of superconductivity is the critical temperature which is the temperature at which the resistance falls to zero – thus as new ways are found to achieve higher critical temperatures, superconductivity is relatively easier to achieve.

## Wind Turbine Energy Generation

$$\frac{peak\_power(W_{peak})}{cost(\$)} = \frac{W_{peak}}{\$}$$

- A measure of the device performance of a wind turbine is the peak power per unit cost. Note that factors that may affect actual energy output such as wind turbine placement are not included in this metric.

## Wireless Information Transmission

$$\frac{kilobits(Kbits)}{time(sec)} = \frac{Kbits}{sec}$$

- Wireless information transmission can be measured by the bandwidth/throughput, which is measured in the amount of information transmitted per unit time. Please note that 1 bit = 8 bytes.

$$\frac{information(bits)}{time(sec) * spectrum(Hz)} = \frac{bits}{sec * Hz}$$ - Wireless information transmission can also be measured by the spectral efficiency, which is a measure of the throughput of the information per wavelength used. Please note that 1 bit = 8 bytes.

$$\frac{information(bits)}{time(sec) * area(m^2)} = \frac{bits}{sec * m^2}$$ - Wireless information transmission can also be measured by the coverage density, which is a measure of the throughput of the information per area covered. Please note that 1 bit = 8 bytes.

Amaya (2008) provides a useful overview of the three wireless information transmission metrics:

*We now proceed to elaborate on the importance that these three aspects of wireless information transportation have for the analysis performed in the present paper. The basic function being studied is transport of information. The FPMs of importance consider the performance relative to some key resource and thus explore engineering tradeoffs over time (ref 1 and 2). In this sense, in the first instance, throughput is of critical relevance for assessing wireless technological progress as time is always a relevant resource. Moreover, the air interface has been generally regarded as a very hostile mean for wireless data transmission. In this way, the possibility of accomplishing high rates of progress in throughput despite the hostility of the transmission environment would provide strong evidence of the capacity of technology to overcome these adverse environmental conditions, made possible through the usage of science in the development of increasingly advanced technologies. This would also represent a clear sign of the new business opportunities that may emerge resulting from such technological progress. Secondly, the transportation efficiency in wireless provides an excellent perspective on the ability of the technology to make better usage of the limited resources contained in the radio spectrum. Radio spectrum is the single scarcest resource in the wireless telecommunications industry and it is therefore particularly relevant to explore the technological ability to transport increasingly larger amounts of information over this limited resource. For the* *remainder of the present paper, this aspect will be called spectral efficiency, which is the name most commonly utilized in the industry. And thirdly, using a measurement of coverage provides an appropriate indication of the ability of the technology to transport large amounts of information to an increasingly higher number of people living in increasingly distributed areas. (Amaya, 2008)*

### 4.1.2. Technological Improvement Rates (TIRs)

As was detailed in section 3.2 of the thesis, the TIRs for each of the DMP's illustrate the improvement of the respective FPM over time. Each of the performance metrics from the preceding sub-section was analyzed and a TIR was extracted for each one of them. For each of the DMPs, the data set was analyzed 4 times to extract TIR for dominated vs non-dominated data points and for the entire date range and just points since 1970 - the summary of the 4 different sub-sets for each DMP is shown in table 13.

| | **All Data Points** | **Non-Dominated** |
|---|---|---|
| **All Year Ranges** | AllAll | AllND |
| **1970+** | 1970All | 1970ND |

**Table 13: Four different subsets of data points for each domain metric pair that were used to calculate the technological improvement rates**

Therefore, including the 4 different subsets of data for the 72 domain-metric-pairs, 274 TIRs were calculated. There were 14 TIRs that were not calculated because when the sub-sets were created 0 data points were left, all but one of which were in the Combustion Engine, Electrical Energy Transmission and Incandescent Illumination domains.

One aspect to note is that the variation *within* the technological domains is less than the variation *between* the TDs. This was noticed with the manufacturing domains in section 3.3 and is shown again in figure 53 where the blue bars represent the average TIR for a particular TD and the smaller error bars represent the standard deviation of the TIR for the ALLND data sets. This fact is important to note, that even in a TD with many different FPMs, a single metric of the TIR can be used to represent the technological improvement of the domain.

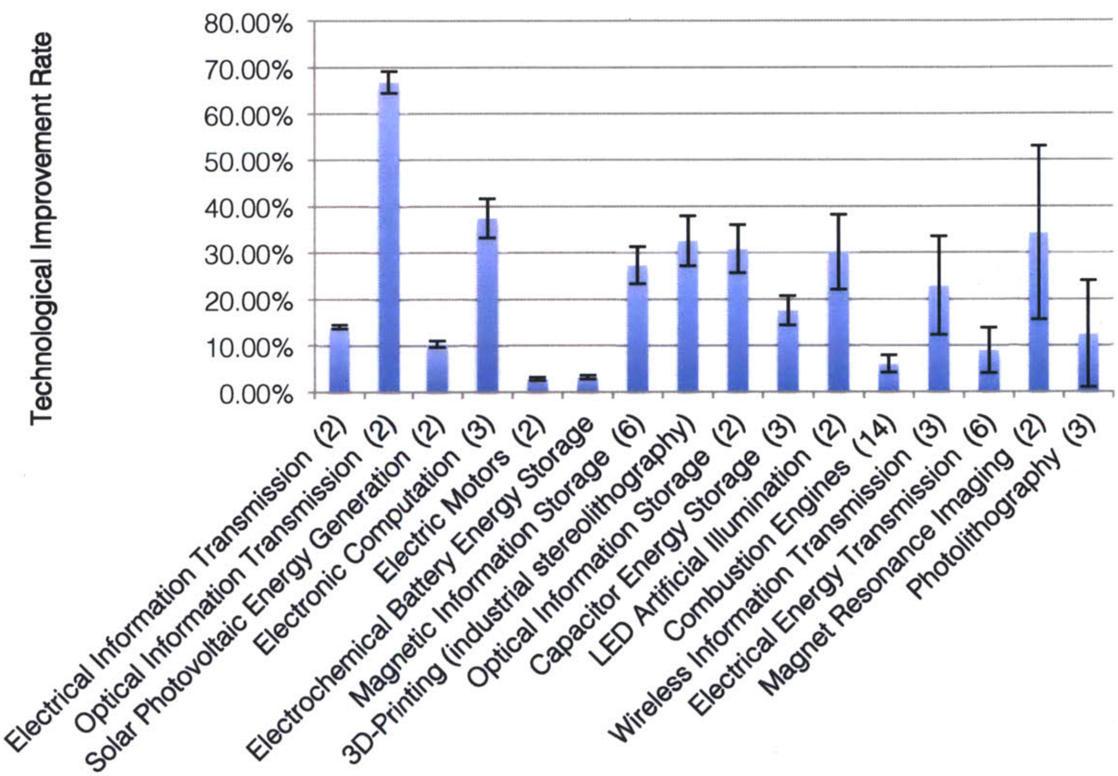



**Figure 53: Mean (blue bars), Standard Deviation (error bars) and number of DMPs (in parentheses on the horizontal axis) of the technological improvement rates for the 16 TDS with multiple FPMs for the AllND data sets (there are 12 TDs with only one FPM)**

It is important to note that in some cases (such as 3D printing, very few data points existed before 1970 and thus the ALL and 1970+ data sets are identical. Other sets, such as electrical information storage, show a large variation in the all and post-1970 data sets. While this is important to note, this is often a case of a significant reduction in the number of data points and thus the values are not reliable. Finally, the difference between the dominated (all data point) sets and the non-dominated sets also shows very little variation. The next sub-section will explore the statistical tests for each of the TIRs.

#### 4.1.2.1. Statistical Overview of the Improvement Rates

Each of the 274 TIRs was subjected to the set of statistical tests described in section 3.2.4 in order to understand the reliability of the data. This section will describe the results of these statistical tests.

The first test that was run is to simply count the number of data points in each of the data sets. In general a higher number of data points indicates the potential for a more reliable metric, although the reliability of a data set is determined also using the other statistical measures in this section. Figure 54 shows the distribution of number of data points for each of the TIRs.

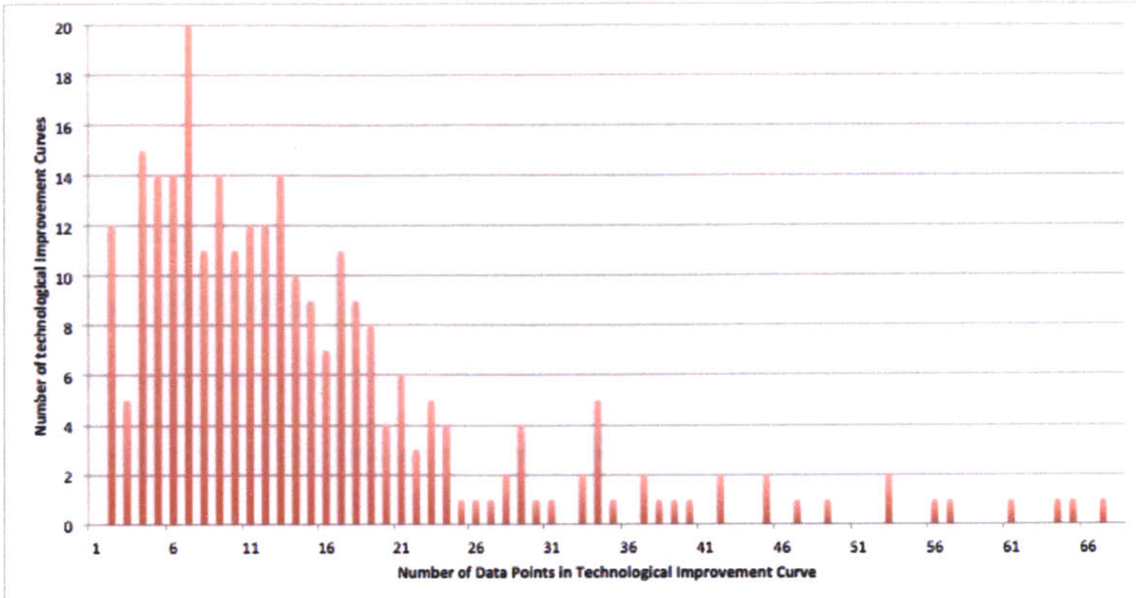



**Figure 54: Distribution of the number of data points for each of the 274 TIRs for the 72 DMPs (5 TIRs with more than 66 data points were removed)**

The number of data points varies significantly from 127 points for the AllAll Combustion Engine (W/L) (not shown) DMP to 2 data points for 15 TIRs. The mean of the long tailed distribution is 15.8 data points with a median of 12.

While the number of data points is a useful measure to get a general idea of the reliability of a TIR, the most common method of evaluating the fit of an exponential regression is the $R^2$ value. Figure 55 shows the distribution of $R^2$ values for each of the 274 TIRs.

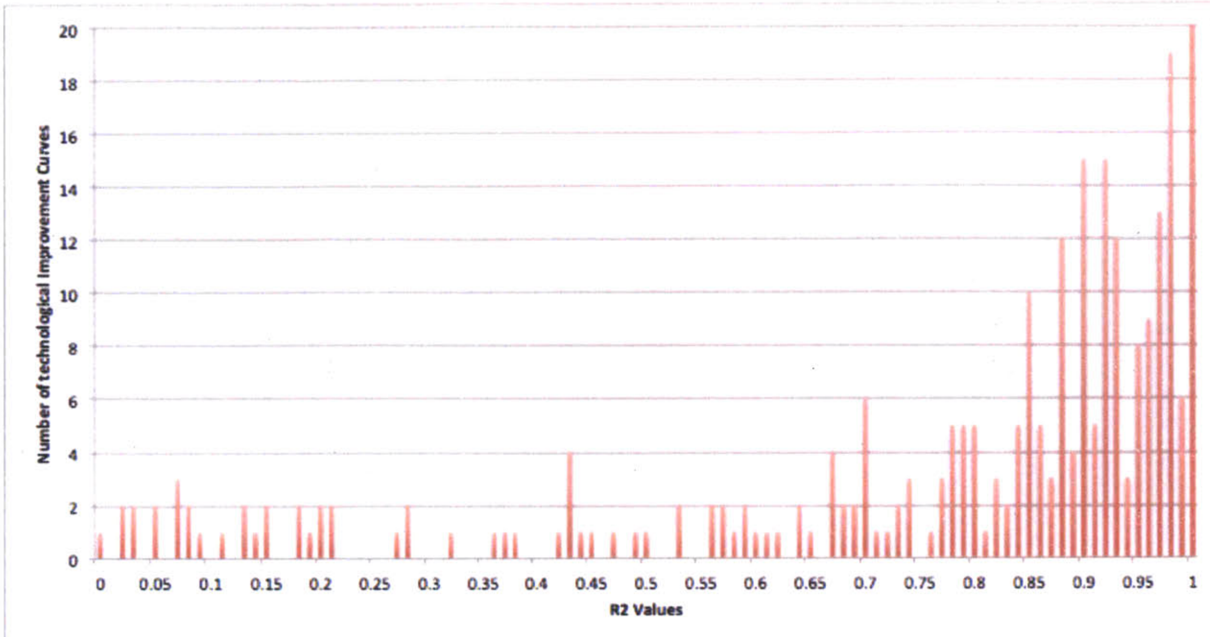



**Figure 55: Distribution of $R^2$ values for the 274 TIRs**

The results in Figure 54 show a large portion (82%) of $R^2$ values above the 0.6 threshold with a mean $R^2$ value of 0.78.

There are some TIR that fall quite short and do not show decent $R^2$ values, even some with a very large number of points, such as the Combustion Engines (W/L) AllAll, which contains 127 data points, yet has only a 0.05 $R^2$. Another examples of a particularly low $R^2$ is Photolithography areal throughput per dollar (70ALL), the low R2 of 0.01 is likely due to the low number of data points present in the data (n=4) as well as it low slope. It is important to note that there are 30 examples of $R^2$ values equal to 1, these are due to a very low number of data points and rounding, for example, any TIR with only 2 data points will necessarily have an $R^2$ of 1.

Another statistical parameter that can be used to evaluate the reliability of the TIR is the standard deviation of the regression - which was explained section 3.2.4.4 of this thesis. Figure 56 shows the variation of the regression standard deviations for the TIRs.

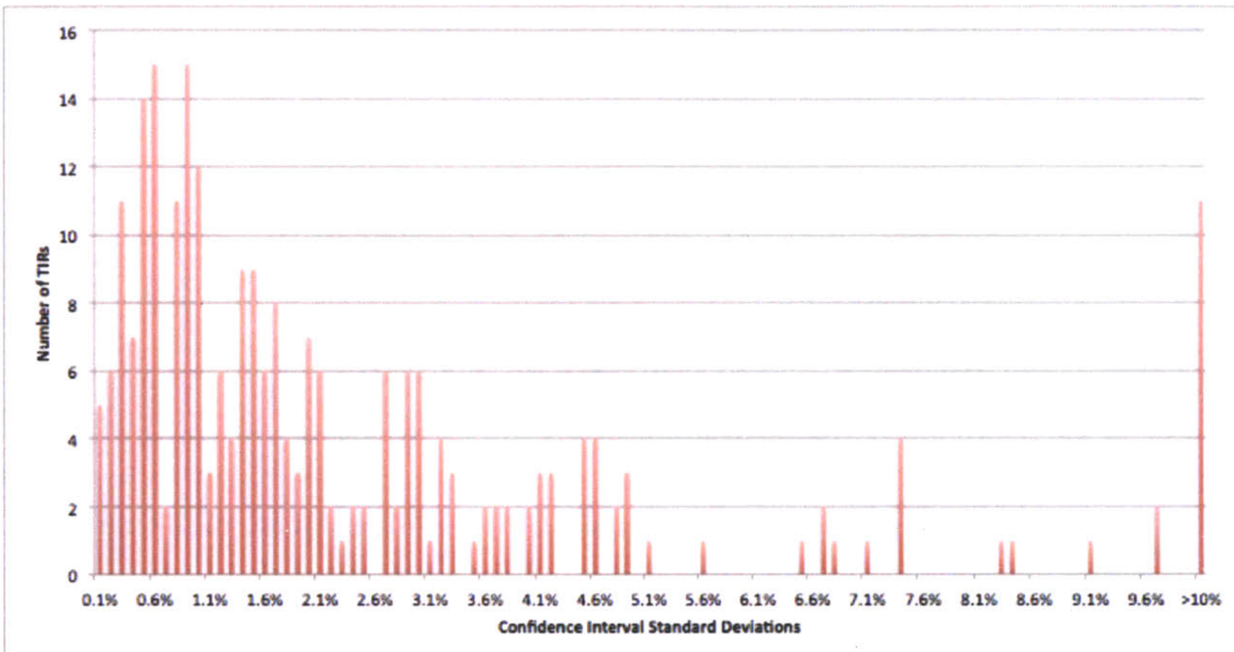



**Figure 56: Distribution of Regression Standard Deviation values for the 274 TIRs**

Most of the regression standard deviations are very small in value with a mean of 1.7%, which is consistent with the $R^2$ data shown above. There are 11 TIRs with very high standard deviations including the 3D Printing metric for $\frac{speed(\frac{mm}{\sec})}{layerthickness(mm)}$ for the non-dominated points that shows a standard deviation of nearly 40% (higher than the TIR). This is likely once again to be due to a limited number of data points (n=3 in this case). The other example to the right in the figure is Optical Information Storage (Mb/$), which only has 4 data points.

As a further test of robustness, each of the data points was systematically removed from each of the TIRs one at a time using the PRM and the resulting TIRs were calculated along with the PRM Standard Deviation (see section 3.2.4.5 for details). This particular metric is used to simulate the effect of adding or subtracting data points to a data set; in general a lower PRM Standard Deviation indicates a more reliable metric. Figure 57 shows the summary of the PRM standard deviations for the TIRs.

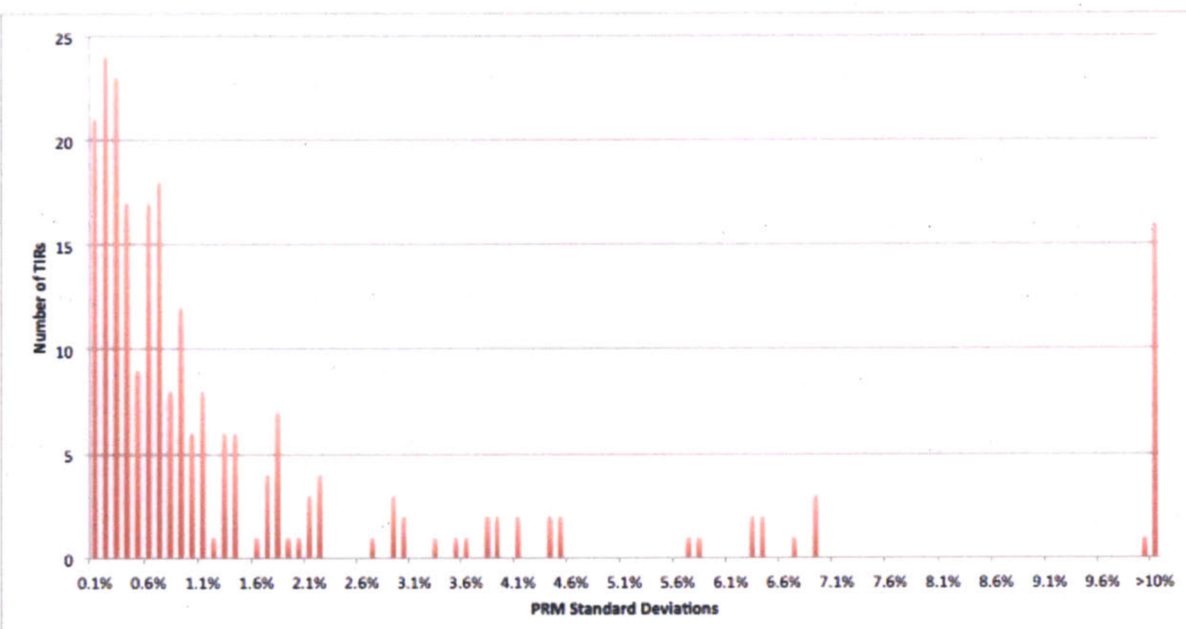



**Figure 57: Distribution of Point Removal Method Standard Deviation values for the 274 TIRs**

The PRM standard deviation values show a low mean of 1.8% with 16 TIRs over 10%. Many of the same unreliable TIRs that were shown using the regression standard deviation are also shown to be unreliable using the PRM. However, there are a few other TIRs that should be analyzed more deeply due to this analysis. The Electronic Computation (1/sec) 70ND data set as well as the MRI (1/mm*sec*$) data sets (all 4!) have a relatively large PRM standard deviation. Once again the cause of this unreliability is likely small amounts of data, but TIRs such as these are removed from the TIR/DPM comparison when possible due their increased sensitivity to the addition and removal of data points. The following section will discuss the selection of the final TIRs for each TD to be used in the cross-domain comparison of the TIRs and the DPMs.

#### 4.1.2.2. Selection of Final TIRs for Each Domain

After all of the analysis, 72 DMP were created and 274 TIR were created (4 sub-sets for each DMP that could be calculated), the intra-domain rates vary considerably less than the inter-domain rates as was demonstrated for manufacturing alone in section 3.3 of this thesis, therefore a <u>Prime</u> TIR is selected to represent each TD in the comparison with the DPMs that are derived for each TD. This sub-section will show the results of this down-selection process using the reliability data gathered in the prior sub-section and the completeness of the metric as was described previously in the thesis.

The main process behind the down-selection is to select the most complete metric that is also reliable. Therefore the process involved selecting the most reliable TIR from the set of 288 and then selecting the most complete TIR for each domain from that set.

As was introduced in section 3.2.4 and shown in the last section, there are a number of ways to locate unreliable data sets, when a domain did not show a high R2, or showed a very low regression-standard deviation or PRM-standard deviation they were considered unreliable and removed. In the same spirit, the TIR that included many data points and had relatively high R2 and low standard deviations (regression and PRM) were considered to be highly reliable data sets. Table 14 shows the down-selection to the 43 most reliable TIRs.

| Domain Name | Prime K | Set | N | R2 | Stdev | PRMstdev |
|---|---|---|---|---|---|---|
| 3D-Printing (industrial stereolithography) (1/sec*$) | 27.30% | 70ND | 4 | 0.97 | 0.049 | 0.045 |
| **3D-Printing (industrial stereolithography) (1/sec*$(including build volume/machine size)))** | **37.60%** | **70All** | **5** | **0.92** | **0.067** | **0.063** |
| **Aircraft Transport (passenger*mph)** | **12.20%** | **AllND** | **12** | **0.98** | **0.006** | **0.002** |
| **Camera Sensitivity (mV/micron2)** | **15.60%** | **70ND** | **11** | **0.99** | **0.005** | **0.002** |
| Capacitor Energy Storage (W*hr/$) | 21.10% | 70ND | 6 | 0.97 | 0.02 | 0.013 |
| **Capacitor Energy Storage (W*hr/kg)** | **14.60%** | **70ND** | **9** | **0.97** | **0.01** | **0.004** |
| Combustion Engines - Aircraft (W/kg) | 3.10% | AllAll | 42 | 0.85 | 0.002 | 0 |
| Combustion Engines - Aircraft Piston (W/kg) | 6.80% | AllAll | 21 | 0.8 | 0.008 | 0.002 |
| Combustion Engines - Aircraft Turbine (W/kg) | 1.70% | 70ND | 8 | 0.98 | 0.001 | 0 |
| Combustion Engines - Automobile (W/$) | 7.90% | AllND | 18 | 0.79 | 0.01 | 0.01 |
| Combustion Engines - Automobile (W/kg) | 5.60% | AllND | 17 | 0.85 | 0.006 | 0.003 |
| **Combustion Engines (W/kg)** | **5.70%** | **AllND** | **24** | **0.82** | **0.006** | **0.002** |
| **Computed Tomography (CT) (1/mm*sec)** | **36.70%** | **70ND** | **13** | **0.78** | **0.045** | **0.018** |
| **Electric Motors (W/kg)** | **3.10%** | **AllND** | **11** | **0.84** | **0.005** | **0.002** |
| Electrical Energy Transmission - AC (W*km) | 14.90% | AllND | 6 | 0.92 | 0.025 | 0.02 |
| **Electrical Energy Transmission (W*km)** | **14.90%** | **AllND** | **10** | **0.97** | **0.009** | **0.004** |
| Electrical Energy Transmission (W*km/$) | 3.60% | AllND | 7 | 0.71 | 0.012 | 0.006 |
| **Electrical Information Transmission (kB/sec*$)** | **14.30%** | **AllND** | **10** | **0.9** | **0.017** | **0.009** |
| Electrochemical Battery Energy Storage (W*hr/$) | 5.70% | 70ND | 9 | 0.91 | 0.008 | 0.004 |
| **Electrochemical Battery Energy Storage (W*hr/kg)** | **7.00%** | **70ND** | **14** | **0.95** | **0.005** | **0.001** |
| **Electronic Computation (1/sec)** | **33.00%** | **AllND** | **19** | **0.9** | **0.02** | **0.007** |
| Electronic Computation (1/sec*$) | 49.10% | 70ND | 18 | 0.93 | 0.021 | 0.011 |
| **Flywheel Energy Storage (kWhr/kg)** | **9.00%** | **70ND** | **7** | **0.92** | **0.014** | **0.011** |
| **Fuel Cell Energy Production (kW/$)** | **14.40%** | **70ND** | **5** | **0.99** | **0.009** | **0.006** |
| **Genome Sequencing (1/$)** | **29.30%** | **70ND** | **7** | **0.91** | **0.038** | **0.038** |
| **Incandescent Artificial Illumination (1000*lm*hr/$)** | **4.50%** | **AllND** | **9** | **0.93** | **0.005** | **0.003** |
| **Integrated Circuit Information Storage (transistors/die)** | **43.20%** | **AllND** | **20** | **0.98** | **0.009** | **0.003** |
| **Integrated Circuit Processors (transistors/die)** | **36.30%** | **AllND** | **12** | **0.97** | **0.016** | **0.007** |
| LED Artificial Illumination (lm) | 24.40% | 70ND | 13 | 0.98 | 0.80% | 0.30% |
| **LED Artificial Illumination (lm/$)** | **36.20%** | **70ND** | **15** | **0.97** | **1.20%** | **0.60%** |
| **Magnet Resonance Imaging (MRI) (1/mm*sec*$)** | **47.50%** | **70ND** | **6** | **0.88** | **0.074** | **0.113** |
| Magnetic Information Storage - Hard Disk (megabits/cc) | 36.10% | 70ND | 19 | 0.94 | 0.017 | 0.007 |
| Magnetic Information Storage - Tape (megabits/cc) | 29.60% | 70ND | 9 | 0.9 | 0.032 | 0.009 |
| **Magnetic Information Storage (megabits/cc)** | **31.90%** | **70ND** | **23** | **0.88** | **0.02** | **0.008** |
| **Milling Machines (hp/mm)** | **3.40%** | **AllND** | **6** | **0.96** | **0.004** | **0.036** |
| **Optical Information Storage (Mb/cc)** | **27.10%** | **70ND** | **15** | **0.95** | **0.015** | **0.007** |
| **Optical Information Transmission (Kb/sec*km*$)** | **65.10%** | **70ND** | **13** | **0.93** | **0.03** | **0.021** |
| **Photolithography (in2/hr*micron)** | **24.00%** | **AllAll** | **11** | **0.85** | **0.03** | **0.008** |
| **Solar Photovoltaic Energy Generation (W/$)** | **9%** | **AllAll** | **35** | **0.94** | **0.04** | **0.01** |
| **Superconductivity (1/K)** | **9.50%** | **70ND** | **7** | **0.73** | **2.90%** | **1.40%** |
| **Wind Turbine Energy Generation (W/$)** | **9.20%** | **70ND** | **8** | **0.93** | **1.10%** | **0.50%** |
| **Wireless Information Transmission (kb/sec)** | **50.40%** | **70ND** | **15** | **0.86** | **0.037** | **0.009** |

**Table 14: The 42 most reliable domain-metric-pairs with TIRs accompanying reliability metrics with final 28 TIRs in BOLD**

Once the set of most reliable data sets were selected, the most complete FPMs for each TD was selected as the representative of for the TIR/DPM correlation comparison. The completeness rankings are shown in the ordering of the DMPs and are shown in a previous subsection 'Domains and FPMs.' When possible, the FPMs selected did not include price parameters due to the fact that their are many outside factors that influence cost in each of these domains and the main focus of this research is on the *technical* improvement rather than the effects of any government intervention. Sometimes (as in the case of Wind Turbines and solar PV), this was not possible, but this fact has been stated several times throughout the thesis and it is important to continue to be aware of.

In some of the cases, there are multiple sub-domains represented (Combustion Engine, Electric Energy Transmission). In these cases, the highest level domain was selected that would include all of the sub-sets. For example, the Combustion Engine data includes all of the combustion engine data points from automobiles and aircraft.

The end result of selection the most complete, reliable TIR is shown in figure 58 for each of the 28 domains.

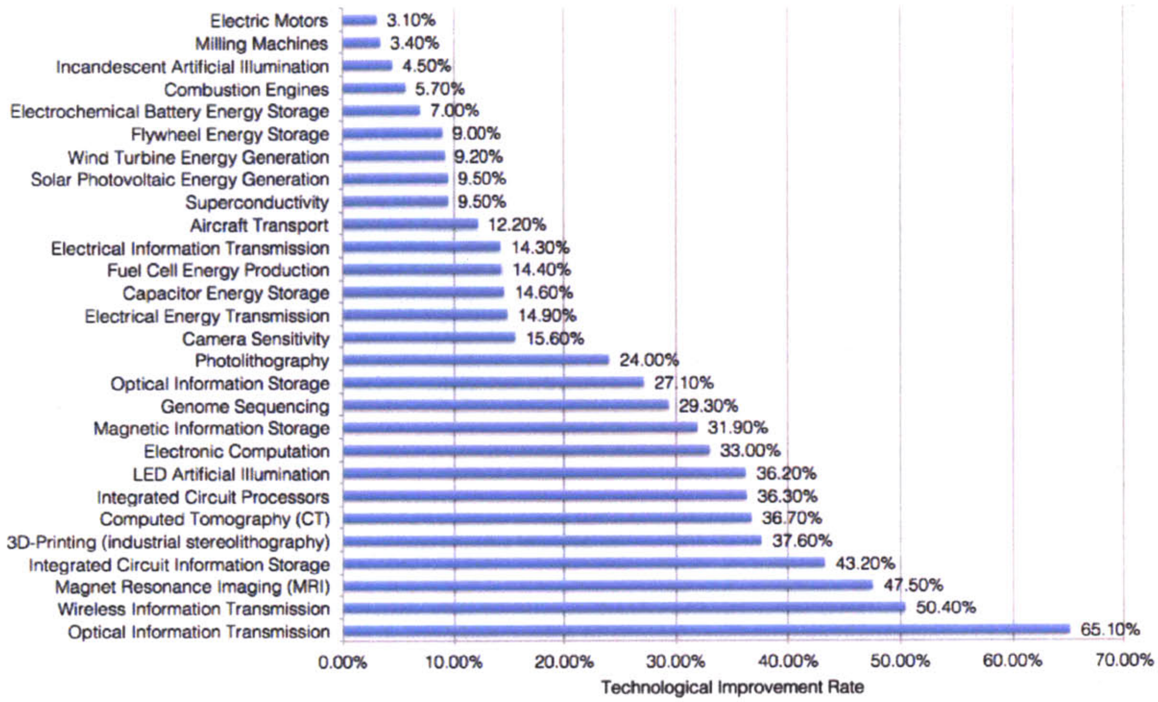



**Figure 58: The most complete and reliable TIR for each of the 28 Domains**

The next section will describe the process of finding the patent sets that represent each of the TDs.

## 4.2. Patent Set Selection: Broad Applicability of the COM

In order to complete the experiment as described, it is necessary to locate relatively complete and relevant patents sets for the 28 different technological domains. This was accomplished by use of the COM, which was thereby shown to be applicable across a wide variety of different technical areas and hierarchy levels; the COM is described in detail in section 3.4. Appropriate patent sets were found for one half of the 28 domains by the direct COM using the overlap of the highest precision and recall UPC and IPC classes. Patent sets for another 8 domains were located with the COM but using the overlap between multiple UPC and IPC classification codes. Finally, 6 of the domains used more intricate variations of the COM to locate the final patent sets.

In this thesis, 28 domains were identified using the COM, there are certainly many more domains that could be classified using this methodology. The total number of patents (including some duplicated) in all of the TDs studied in this thesis is 511,247 and the number of cited patents analyzed was 2,619,355, which is a non-trivial portion of the 4,666,574 patents that were issued between 1976 and 2013 (uspto.gov, 2014). This places an upper limit of 10.9% on the percentage of patents that have been categorized into TDs, and 56.1% on the percentage of cited patents that have been analyzed. Therefore it is likely that less than 10% of the technological domains have been categorized and unlike other categorization methods, the COM allows repeats, which could potentially allow for an incredibly large number of potential TDs that vary in size and scope. Realistically, the number of technological domains that would comprise nearly all of the patents could be in the range of 300-1000 TDs based upon the number of patents and domains that were analyzed in this study.

Another way of examining the total number of domains is to look at where the knowledge comes from, which can be measured by the backward citations of the patents. Using this method, nearly half of the knowledge base since 1976 has been used by the patents in the 28 domains in this thesis. This indicates that although only 10% of the total domains have been analyzed, nearly half of the knowledge base was covered in this experiment, demonstrating the breadth of the experiment. In all likelihood, our TDs cover between 10% and 56% of the total knowledge base in the patent system.

### 4.2.1. Pre-Search Variables

As was mentioned in section 3.4.1, in order to locate the patent set for each TD, a set of search terms was developed that represented the domain and to find a starting patent set that was then used as the input for the COM. These terms were selected based upon prior knowledge of the domain, through talking with experts in the field, and through literature searches including Wikipedia where a number of synonyms are often listed for each domain. For example, the search terms used for LED lighting are as follows: LED, light emitting diode, semiconductor light, electroluminescent, diode lamp, solid state light, solid state illumination. The list of the pre-search keywords for each domain is shown in appendix B of this thesis. Most of the TD's have a significant number of keyword searches; this is done intentionally to provide a wide search breadth initially to help ensure that the most complete and most precise patent set is selected.

Following the COM process, each of the terms was used to search in the title or abstract of all US issued patents since 1976 and a primary UPC and IPC for each search term were designated based upon their calculated MPR scores. The MPR scores, which are covered in section 4.2.2 below are a measure of how closely related the UPC or IPC is with a particular search term, and thus higher MPRs are better. Additionally, the overall size of the pre-search using the search term is included for each of the domains to help assess completeness of the patent sets eventually selected.

Nearly all of the domains were narrowed down this way, with the few exceptions that will be mentioned at the end of this section. The distribution of the MPR values of all of the keyword pre-searches for the 28 TDs is shown in Figure 59.

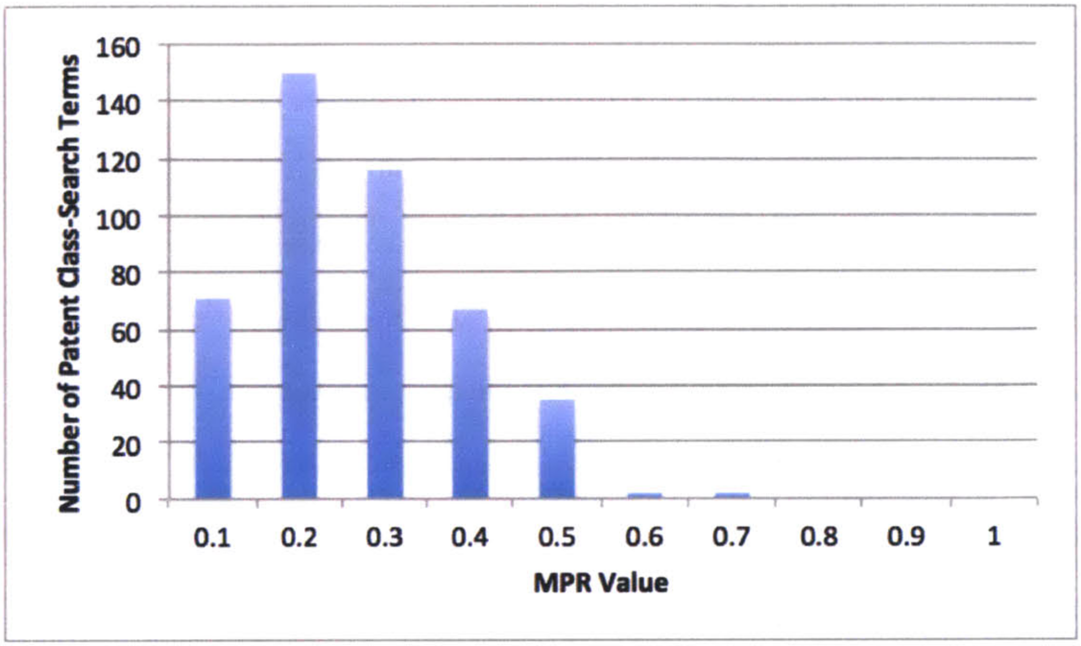



**Figure 59: Distribution of mean-precision-recall values for Patent Class-Search Term Pairs**

The distribution is skewed towards lower numbers and has a mean value of 0.217 and a standard deviation of 0.117, showing a rather high variability. Very high MPRs are nonexistent which lends support to our earlier observation (Benson and Magee, 2013) that the IPCs and UPCs do not on their own correlate well with specific keyword search terms.

### 4.2.2. Direct COM

As was mentioned earlier, patent sets for 14 of the 28 TD were located using the simple overlap between one IPC and one UPC. This result shows the surprising ease of which highly relevant and complete data sets can be located using the COM. All of the patent sets that were located using the direct COM had relevancy higher that 80%. Table 15 shows a summary of the patent sets selected for the 14 TD using the direct COM method, and more in depth information about the selection of each of the direct COM patent sets can be found in Appendix B.1.

| TD | Size | Relevancy | Patent Class Overlap |
|---|---|---|---|
| Camera Sensitivity | 1744 | 86% | 257 AND H04N |
| Capacitor Energy Storage | 5944 | 84% | 361 AND H01G |
| Electric Motors | 17869 | 86% | 310 AND H02K |
| Electrical Energy Transmission | 10375 | 86% | 363 AND H02M |
| Electrical Information Transmission | 44910 | 67% | 439 AND H01R |
| Electronic Computation | 13204 | 97% | 712 AND G06F |
| Integrated Circuit Information Storage | 49018 | 81% | 365 AND G11C |
| Integrated Circuit Processors | 149491 | 81% | 257 AND H01L |
| LED Artificial Illumination | 3792 | 85% | 313 AND H01L |
| Magnetic Information Storage | 33576 | 93% | 360 AND G11B |
| Milling Machines | 2315 | 93% | 409 AND B23C |
| Optical Information Storage | 23543 | 82% | 369 AND G11B |
| Solar Photovoltaic Energy Generation | 5203 | 85% | 136 AND H01L |
| Superconductivity | 1776 | 85% | 505 AND H01L |

**Table 15: Patent Sets for the 14 domains that were found using the Direct COM including the UPC and IPC class used in the overlap**

## 4.2.3. Multiple UPC or IPC classes used in the COM Overlap

As was discussed briefly in section 4.2.3, the COM can be adapted to use the overlap of more than two patent classifications as long as there is at least one UPC and one IPC (i.e. the overlap between 3 UPCs would not work). Out of the 28 TDs, 8 of the patent sets were located by using the overlaps of 3 or more classifications. The specific details of pre-searches and preliminary patent set relevancy percentages can be found in appendix B.2 of the thesis. The patent sets found using 3 or more classification and the COM are given in table 16.

| TD | Size | Relevancy | Patent Class Overlap |
|---|---|---|---|
| Combustion Engines | 19094 | 96% | 123 AND (F01L OR F02B) |
| Computed Tomography (CT) | 6817 | 88% | 378 AND (A61B OR G01N) |
| Incandescent Artificial Illumination | 642 | 89% | (313 AND H01K) AND NOT (H01J1 OR F21V) |
| Magnet Resonance Imaging (MRI) | 1778 | 86% | (324 AND A61B) OR (600 AND G01R) |
| Optical Information Transmission | 36494 | 82% | (398 AND H04B) OR (385 AND G02B) |
| Photolithography | 14975 | 87% | (430 OR 355) AND G03F |
| Wind Turbine Energy Generation | 2498 | 94% | (416 OR 290) AND F03D |
| Wireless Information Transmission | 39675 | 94% | 455 AND (H01L OR H04B) |

**Table 16: Summary of Patent Sets for the 8 Patent Sets that were found using the COM with overlap of 3+ Patent Classifications including the classifications used in the overlap**

## 4.2.4. COM Modifications

While many of the TDs were relatively easy to find using the COM, there were a few that required deeper searching and more sophisticated applications of the COM. Such methods were described in the methodology section, using the solar PV, wind turbine, batteries and capacitors as the examples, therefore these domains will not be covered. However, there were a few other unique uses of the COM that were utilized in order to locate sets of patents to represent some of the TDs.

The summary of the 6 TDs in which the COM modifications were used is shown in table 17. Further details on the selection of the patent sets can be found in appendix B.3 of this thesis.

| TD | Size | Relevancy | Patent Class Overlap |
|---|---|---|---|
| 3D-Printing (industrial stereolithography) | 251 | 93% | 264/401 AND B29C35/08 |
| Aircraft Transport | 8629 | 79% | 244 AND (B64D OR B64C) AND NOT ('canopy' OR 'parachute' or 'helicopter') |
| Electrochemical Battery Energy Storage | 16122 | 83% | (429 AND H01M) AND NOT 'fuel cell' |
| Flywheel Energy Storage | 154 | 70% | 74/572 AND (F16F15 OR H02K7) |
| Fuel Cell Energy Production | 7368 | 97% | (429 AND H01M) AND 'fuel cell' |
| Genome Sequencing | 3990 | 74% | (435/6.11 OR 435/6.12) AND C12Q |

**Table 17: Patent Sets for the 6 Patent Sets that were found using the COM with Modifications. The classes are also given now, but usually deeper in the patent classification hierarchy.**

#### 4.2.4.1. Lower level Hierarchy Classifications

The COM was designed to work at the primary level of the UPC (before the /) and the 4 digit level of the IPC (H01L). Finding highly relevant patent sets by overlaps of high MPR IPC and UPC classes also works at lower level hierarchy classifications. An example of this is 3D printing, where the primary UPC located is 264 (Plastic and nonmetallic article shaping or treating: processes), however the more appropriate patent class for SLA 3D Printing is 264/401 (STEREOLITHOGRAPHIC SHAPING FROM LIQUID PRECURSOR). This same approach can be applied to the IPCs in SLA 3D printing with the primary IPC being B29C (SHAPING OR JOINING OF PLASTICS; SHAPING OF SUBSTANCES IN A PLASTIC STATE, IN GENERAL; AFTER- TREATMENT OF THE SHAPED PRODUCTS, e.g. REPAIRING) and the appropriate IPC being B29C35/08 (Heating, cooling or curing, e.g. crosslinking, vulcanising; Apparatus therefor... by wave energy or particle radiation). These lower level hierarchy classifications are overlapped in the same way to find the appropriate patent sets.

#### 4.2.4.2. Pre-Searching Using Known Company Names

The pre-search using keywords works very well for most of the domains, and when additional knowledge is known about a particular domain, that knowledge is used to help locate the patent classifications of interest. In particular searching for the patents that are assigned to companies that are known to work in a particular TD can act as a useful supplement to the initial keyword search. This technique was used in selecting the patents for the Genome Sequencing TD, as there were a few well-known organizations that worked on Genome Sequencing (eg. Affymetrix, Oxford Nanopore Sciences, Sequenom, Illumina, Knome, Broad Institute) and thus helped located the final patent classification codes. It should be noted that using only the company names as a pre-search may result in the selection of a patent classification that is more associated with a particular company than a TD.

#### 4.2.4.3. Keyword Cleaning

In the pursuit of very complete and relevant patent sets, in some domains, it is advantageous to remove or add to the final classification overlap patent set by using keywords. This approach, while it can improve the relevance/completeness of a particular patent set, should not be used broadly as it can make the repeatability of locating a patent set more difficult. Three of the 28 TDs used some keyword cleaning by removing patents that are not related to the TD of interest (i.e. Aircraft Transport where 'parachute', 'canopy' and 'helicopter' were removed). An important case is that batteries and fuel cells were actually in the same set of patents (429 AND H01M) but were easily separated by final key word editing (using the keywords 'fuel cells')

### 4.2.5. Selecting the 100 most important patents

Along with the selection of a patent set to represent each TD, the patents were rank ordered by number of forward citations and the top 150 patents were all read and non-relevant patents removed to form a 100% relevant set of the 100 most cited patents within each TD as was discussed in section 3.4.4. The reading of the top 150 patents was also included in the relevancy percentage, therefore the relevancy rankings are weighted towards the more highly cited patents within a domain. Figure 60 shows the overview of how the relevancy scores were determined.

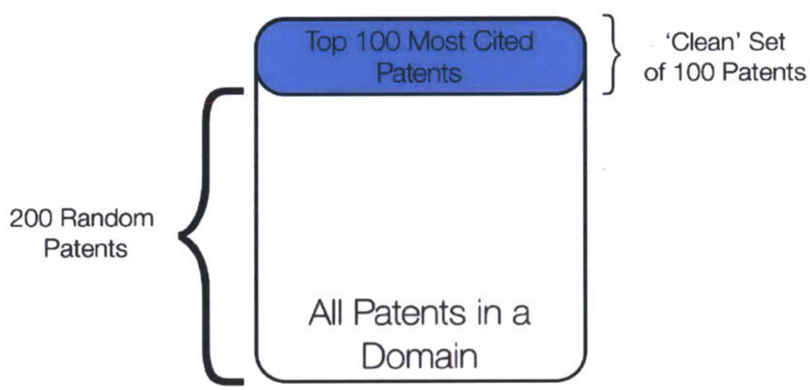



**Figure 60: Graphical Representation of how the Relevancy Rankings and Top 100 Clean Patents were determined**

### 4.2.6. Final Data Sets and Relevancy Scores

Ultimately, patent sets and clean 'Top 100' patents for all 28 technological domains were located and relevancy scores were determined. The summary of the patent data sets is shown in figure 61 on a logarithmic scale.

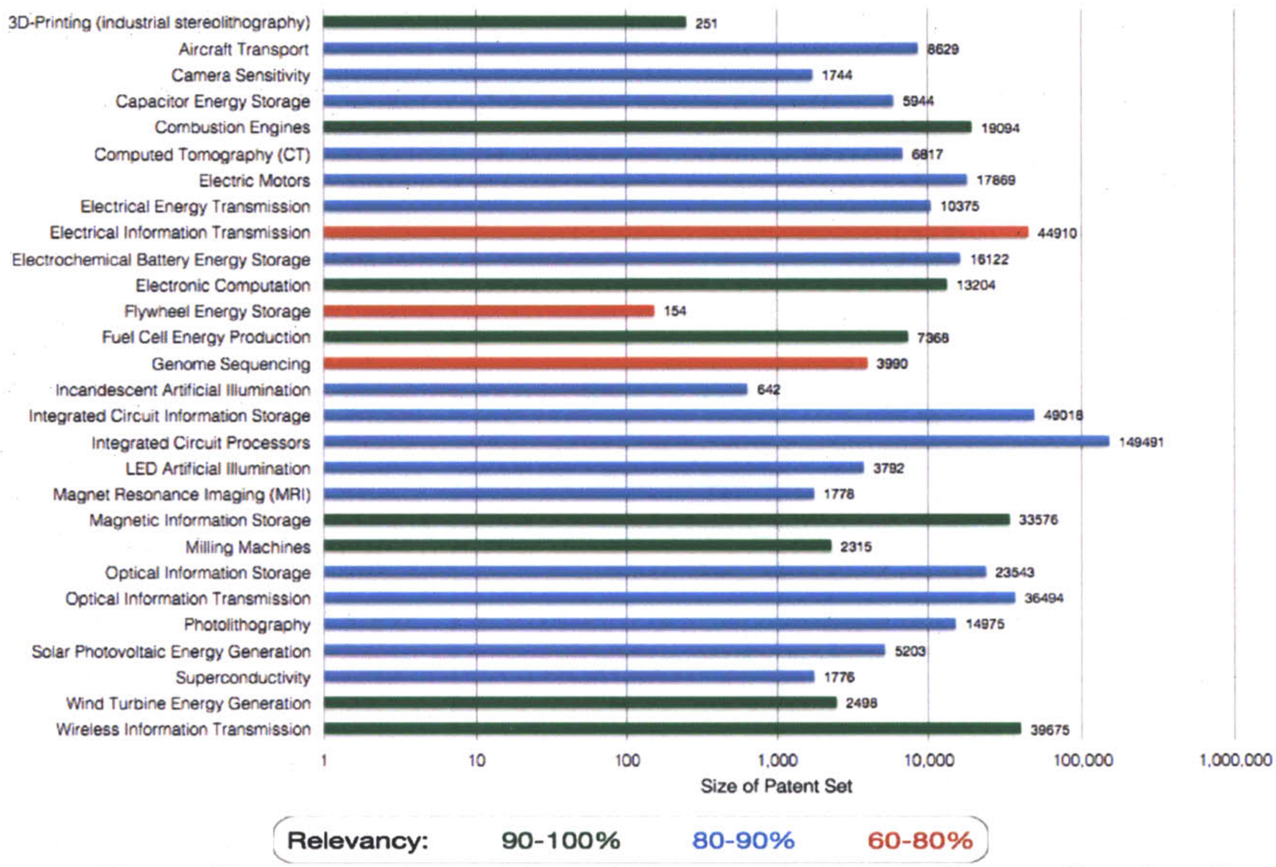



**Figure 61: Size (on log scale) and Relevancy (in color) of Patent Sets for all 28 technological domains**

Nearly all of the patent sets have relevancy percentages higher than 80%, with 3 showing relevancies between 60 and 80%. The overall size of the patent sets ranges from 154 (Flywheel Energy Storage) to 149491 (Integrated Circuit Processors). Each of these patent sets is used to represent their respective TD in the comparison with the TIRs. In addition, the clean top 100 will also be used in some of the comparisons.

### 4.2.7. Most Closely Related Patent Sets

One of the results of locating these patents was the ability to see the overlap between the patents. Because each patent can be multiply listed in a number of different UPCs and IPCs, some patents will be present in multiple patent sets in the patents selected to represent the 28 TDs examined for this research.

In order to quantify the overlap between the patents, each patent set was compared with each of the other 27 domains in order to find the percentage overlap between the patent sets. This ratio is shown in equation 36, with the patent set of interest represented by $P_i$ and each the other 27 patent sets represented by $P_j$.

$$\frac{P_i \cap P_j}{P_i} \qquad \text{(Equation 36)}$$

For example, there are only three patents that are present in both the Electrochemical Battery Energy Storage TD and the Aircraft Transport TD, and there are 8629 patents in the Batteries TD, therefore the overlap of Aircraft with Batteries is 0.000347 as is shown below.

$$\frac{P_{batteries} \bigcap P_{aircraft}}{P_{batteries}} = \frac{3}{8629} = 0.000347$$

This very small overlap is not surprising but neither is the fact that many of the patent sets had a significant overlap with Integrated Circuit Processors due to its large size and the relative ubiquity of integrated circuits. Figure 62 shows the overlap of the other domains with Integrated Circuit Processors (the 12 domains with less than 0.01% overlap are not shown).

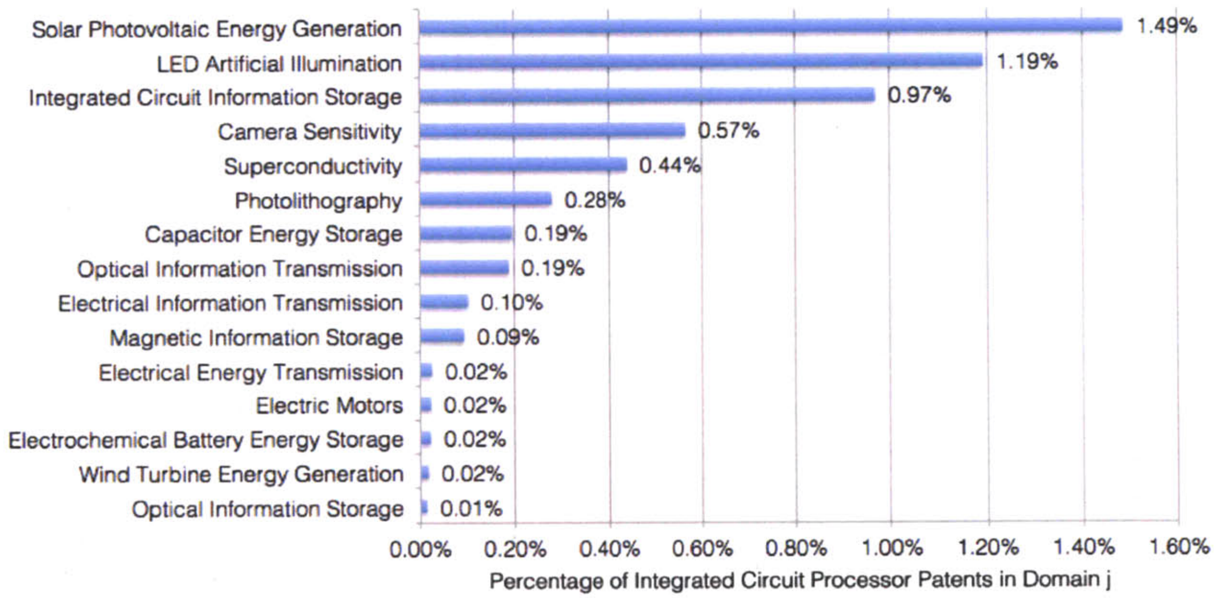



**Figure 62: Overlap of Integrated Circuit Processors Patents with patents in other domains (12 domains with less than 0.01% overlap are not shown).**

There are a number of semiconductor-related TDs that are closely related to IC processors and while the amount of overlap with the large IC class is not significant with the denominator being this large class, this changes when the opposite ratio is formed. For example, while only 1.49% of the IC Processor patents overlap with Solar PV, this number of doubly listed patents is nearly half of the Solar PV patents as is shown in figure 63.

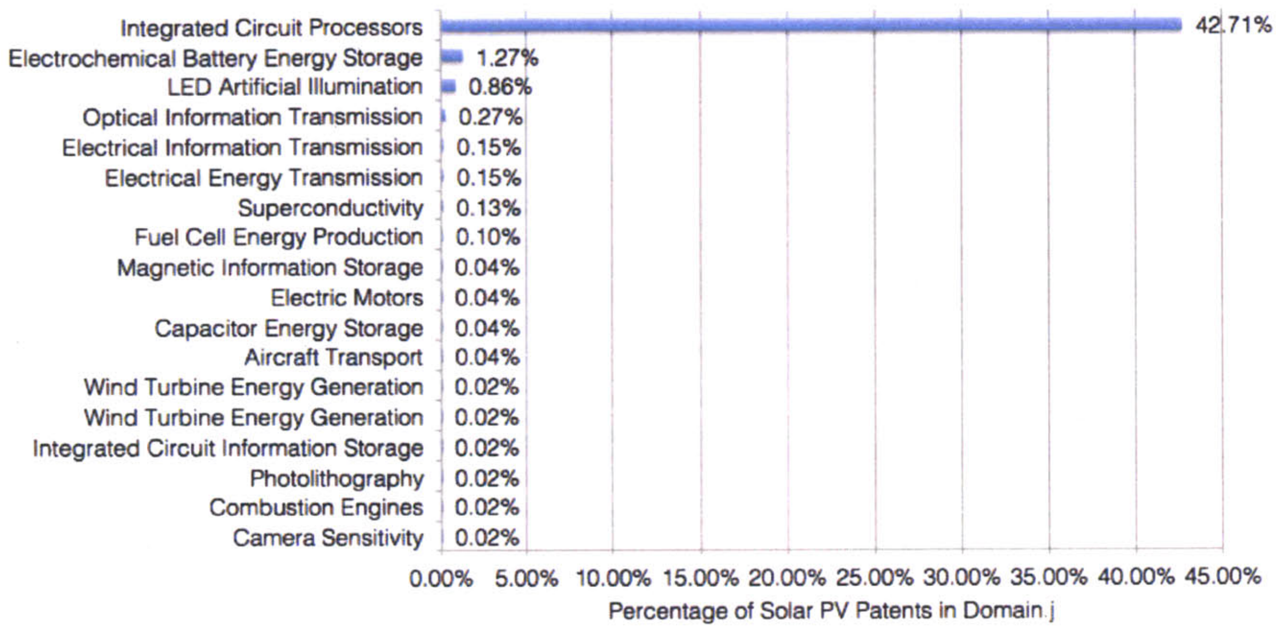



**Figure 63: Overlap of Solar PV patents with patents in other domains (8 domains with less than 0.1% overlap are not shown).**

Using these overlap percentages, we found the patent sets that are most closely related to each other through direct overlap. Table 18 shows the TD (j) with the highest overlap for that particular TD (i) for all 28 domains.

| Patent Sets (i) | Maximum Overlapping Patent Set (j) | Percent Overlap |
|---|---|---|
| Camera Sensitivity | Integrated Circuit Processors | 48.5% |
| LED Artificial Illumination | Integrated Circuit Processors | 47.0% |
| Solar Photovoltaic Energy Generation | Integrated Circuit Processors | 42.7% |
| Superconductivity | Integrated Circuit Processors | 37.0% |
| Flywheel Energy Storage | Electric Motors | 26.0% |
| Fuel Cell Energy Production | Electrochemical Battery Energy Storage | 14.9% |
| Optical Information Storage | Magnetic Information Storage | 13.5% |
| Magnetic Information Storage | Optical Information Storage | 9.5% |
| 3D-Printing (industrial stereolithography) | Photolithography | 7.6% |

| | | |
|---|---|---|
| Electrochemical Battery Energy Storage | Fuel Cell Energy Production | 6.8% |
| Capacitor Energy Storage | Integrated Circuit Processors | 4.9% |
| Integrated Circuit Information Storage | Integrated Circuit Processors | 3.7% |
| Photolithography | Integrated Circuit Processors | 2.8% |
| Magnet Resonance Imaging (MRI) | Computed Tomography (CT) | 1.7% |
| Integrated Circuit Processors | Solar Photovoltaic Energy Generation | 1.5% |
| Incandescent Artificial Illumination | Electrical Information Transmission | 1.4% |
| Electric Motors | Magnetic Information Storage | 1.3% |
| Wind Turbine Energy Generation | Electric Motors | 1.0% |
| Optical Information Transmission | Wireless Information Transmission | 0.8% |
| Wireless Information Transmission | Optical Information Transmission | 0.7% |
| Computed Tomography (CT) | Magnet Resonance Imaging (MRI) | 0.5% |
| Electrical Energy Transmission | Integrated Circuit Processors | 0.4% |
| Electronic Computation | Integrated Circuit Information Storage | 0.3% |
| Electrical Information Transmission | Integrated Circuit Processors | 0.3% |
| Milling Machines | Electric Motors | 0.3% |
| Aircraft Transport | Wind Turbine Energy Generation | 0.1% |
| Genome Sequencing | Integrated Circuit Processors | 0.1% |
| Combustion Engines | Electric Motors | 0.1% |

**Table 18: Most Closely Related Pairs of Patent Sets showing overlapping patent percentage**

Ten of the TD are most closely related to Integrated Circuit processors, including the 4 highest overlapping sets. There are a few patent sets that are nearly 50% overlapped with another TD, 21 out of the 28 patent sets have less than 10% overlap. This is a very nice quantitative demonstration of the reality of the status of ICs as a general-purpose technology (GPT) (Bresnahan and Trajtenberg, 1995) and even support for the idea of IC's being the most important GPT ever (Brynjollfosson and McAfee, 2011).

## 4.3. Relating Domain Patent Markers to the TIRs

Following the 5 hypotheses described in section 3.5, the domain patent markers are used to test hypotheses derived from theories of technological change. In this section, each of the hypotheses is briefly restated and the relevant DPMs are correlated with the TIRs in order to determine whether the hypotheses are supported by the data. Each DPM is plotted against the TIR and a Pearson correlation coefficient is given along with the result of the test of the rejection of the null hypothesis at a 5% confidence level. If the null hypothesis is accepted, the correlation is likely to be simply a result of the random scattering of the data, whereas if it is rejected, the correlation is likely due to true variation in the parameters being measures. The minimum and maximum values are also given alongside a short analysis of the results as needed. More in depth discussions of the results of the DPM and TIR comparison are included in the Chapter 5 of the thesis.

### 4.3.1. Hypothesis 1: Effort

Hypothesis 1 states that a higher number of patents should relate to the technological improvement rate of a domain as shown diagrammatically in Figure 64.

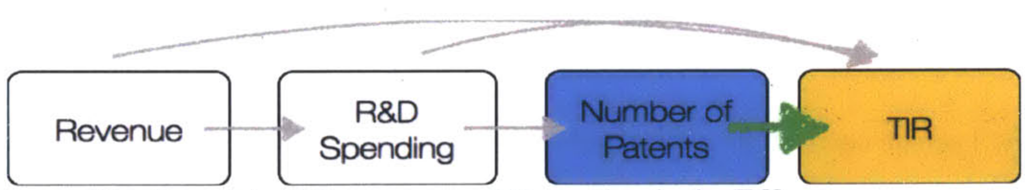



**Figure 64: Graphical Summary of Hypothesis 1 - Effort**

#### 4.3.1.1. Simple Patent Count

The measure used to test this is the simple patent count, which is the number of patents in a particular TD between January 1, 1976 and July 1, 2013. Figure 65 shows the relationship between TIR and simple patent count.

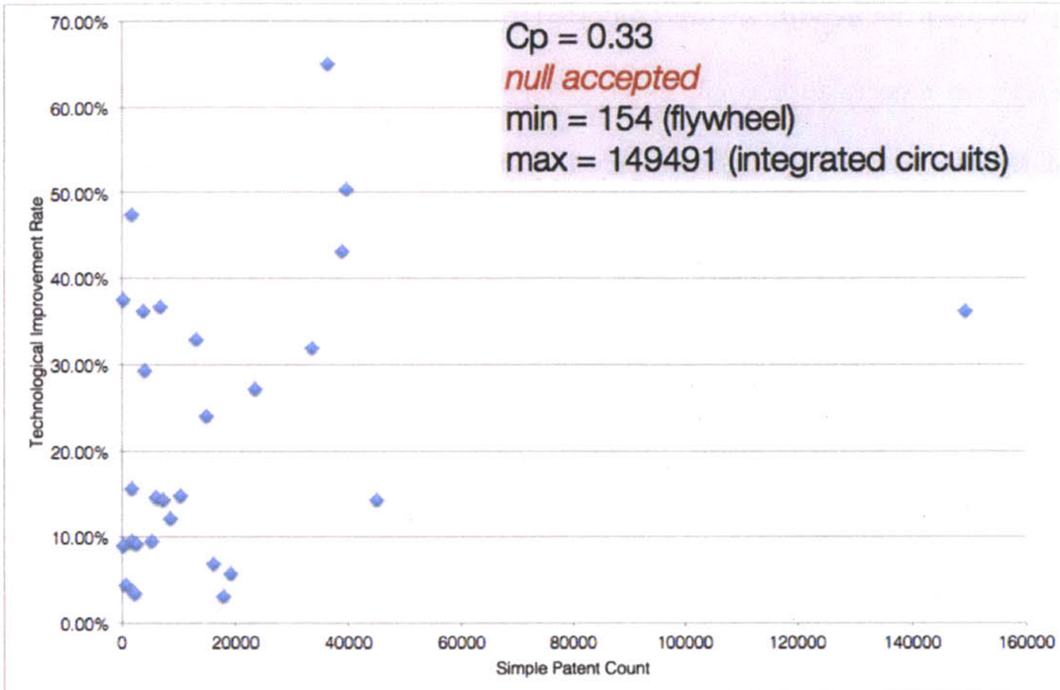



**Figure 65: Technological Improvement Rates vs Simple Patent Count; the Pearson correlation coefficient ($c_p$), the null hypothesis acceptance and the values of the independent variable for the domains having maximum and minimum values are shown in the upper right for this graph and the following results plots through section 4.3.6.**

The Pearson correlation coefficient between the two variables is 0.33, however the null hypothesis that the correlation could be do to the random variation in the data was accepted as is shown above. The combination of the statistical tests and the lack of any discernible trend in figure 65 indicate that there is not a strong relationship between the number of patents in a technological domain and the associated TIR. Thus, in this form H1 is not supported by the data.

### 4.3.2. Hypothesis 2: Breakthroughs

Hypothesis 2 states that only very highly cited patents contribute to the improvement of a TD and therefore a TD with a higher number of highly cited patents should have a higher TIR. This relationship is diagrammed in Figure 66.

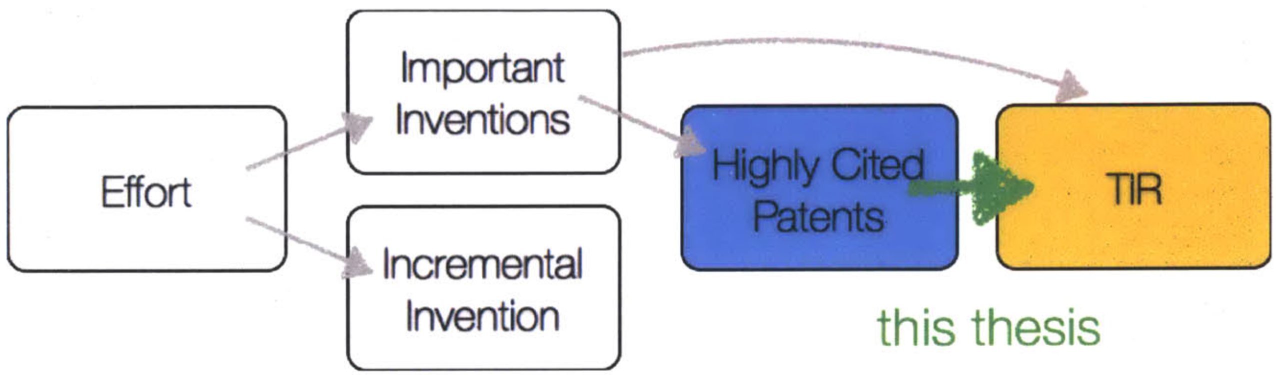



**Figure 66: Graphical Summary of Hypothesis 2 - Breakthroughs**

#### 4.3.2.1. Average Number of Forward Citations

One of the measures used to test hypothesis 2 is the average number of forward citations received by the patents in a TD. Figure 67 shows the relationship between TIR and the average number of forward citations.

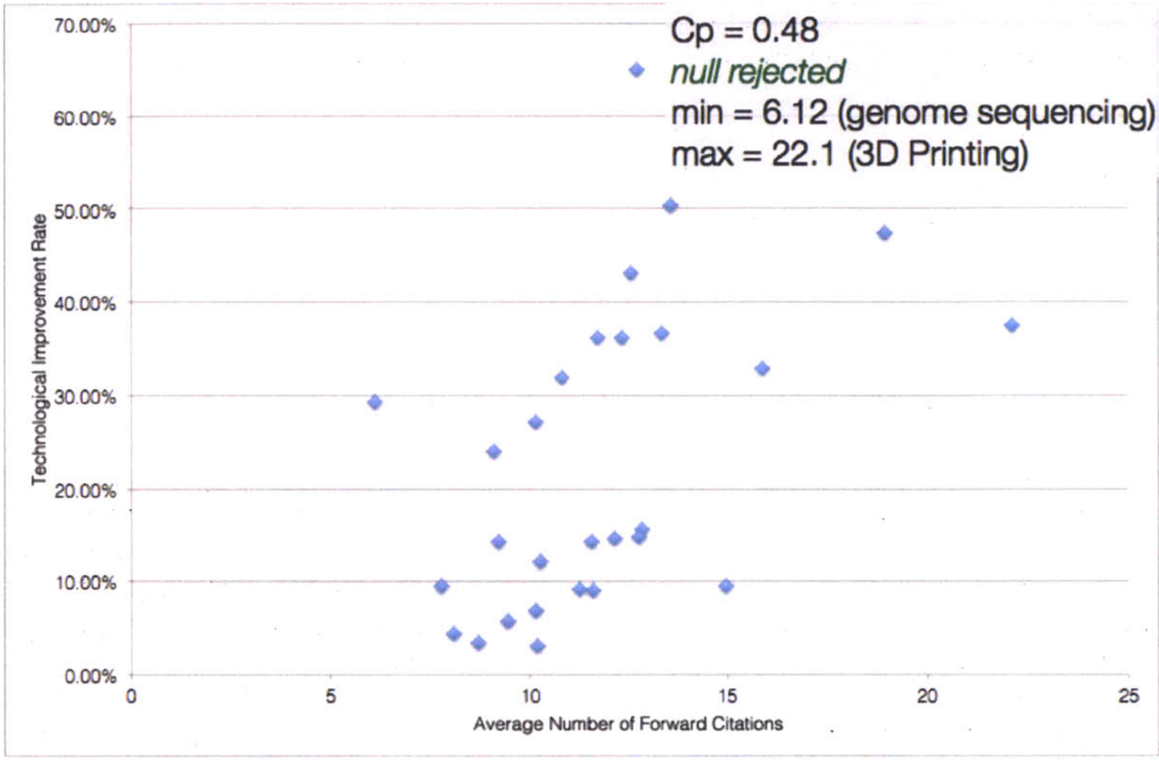



**Figure 67: TIR vs Average Number of Forward Citations**

The Pearson correlation coefficient between the two variables is 0.48, and the null hypothesis is rejected, indicating that the correlation is unlikely due to random scattering of the data. The combination of the statistical tests and the slight visual trend in figure 66 indicate that there is a potential relationship between the average number of forward citations per patent in a technological domain and the associated TIR.

It is interesting to note that the highest number of forward citations (3D Printing) and the lowest (Genome Sequencing) are both relatively rapidly improving domains and the fastest improving domain (Optical Telecom) is near the median of the 28 domains showing overall that the correlation is not at all infallible.

#### 4.3.2.2. Ratio of Patents with more than 20 Forward Citations

Another measure used to test hypothesis 2 is the percentage of patents in a TD that receive a large number (over 20) of forward citations, which is based off of the work of Schoenmakers and Duysters (2010). Figure 68 shows the relationship between TIR and the percentage of patents in a TD with more than 20 forward citations.

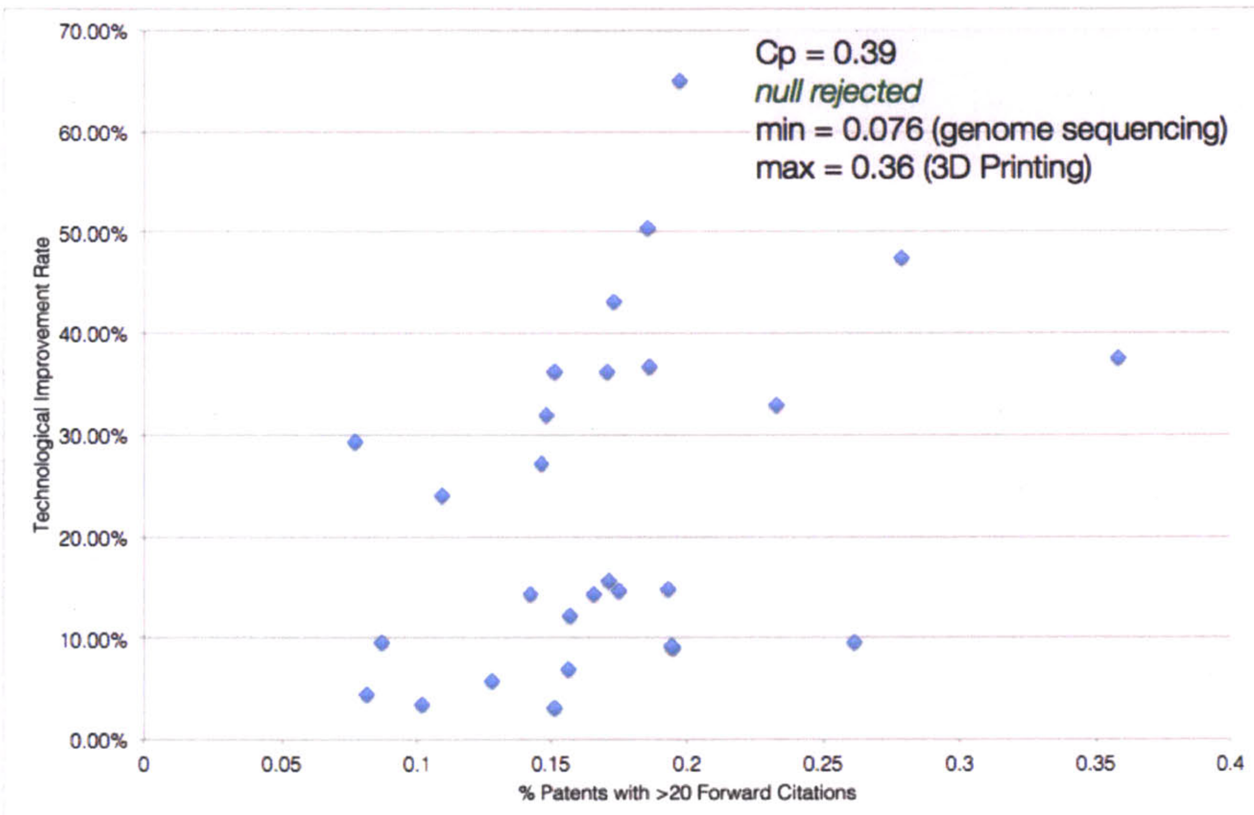



**Figure 68: TIR vs % of Patents with more than 20 Forward Citations**

The Pearson correlation coefficient between the two variables is 0.39, and the null hypothesis is rejected, indicating that the correlation is unlikely due to random scattering of the data. The combination of the statistical tests and the slight trend in figure 68 indicate that there is a potential relationship between the average number of forward citations per patent in a technological domain and the associated TIR.

The distribution of the domains using this metric is very similar to that of the average number of forward citations, and this close relationship is further supported by the fact that the cross-correlation between the two metrics is a very high 0.96 (rejected null hypothesis). This extremely high correlation indicates that the two measures are quite similar. Overall, the tests indicate empirical support for Hypothesis 2.

### 4.3.3. Hypothesis 3: Science

Hypothesis 3 is based upon the idea that basic science is one of the main drivers of the improvement of technology. The Hypothesis states that domains with higher average non-patent citations will have higher TIR, as is diagrammed in Figure 69.

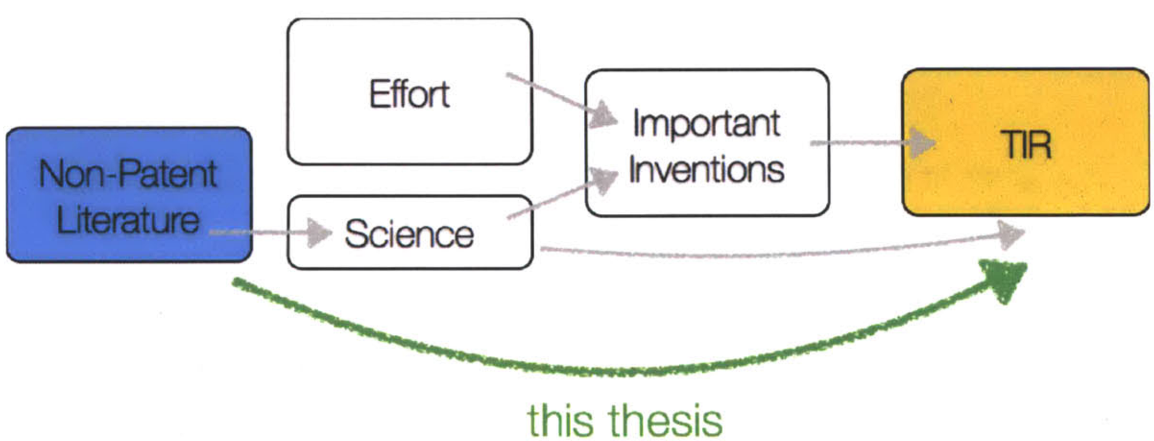



**Figure 69: Graphical Summary of Hypothesis 3 - NPL**

Figure 70 shows the relationship between TIR and the average non-patent literature citation ratio for the patents in a TD.

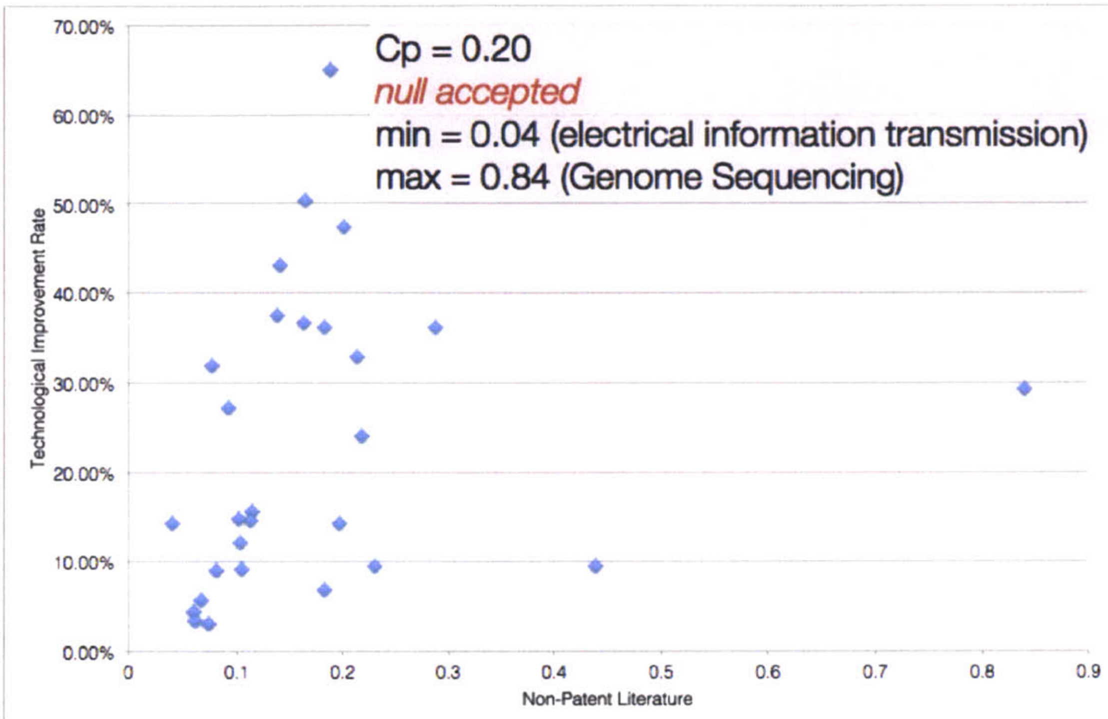



**Figure 70: TIR vs Average Non-Patent Literature Citation Ratio**

The Pearson correlation coefficient between the two variables is 0.20, and the null hypothesis is accepted, indicating that the slight correlation could be due to random scattering of the data. The combination of the statistical tests and no obvious trend in figure 70 indicate that there is not a strong relationship between the average NPL citation ratio in a technological domain and the associated TIR.

The very large value of the NPL ratio for genome sequencing is of note, in that nearly all of the citations that come from genome sequencing patents are to scientific literature, nearly double the 2nd highest TD (superconductivity NPL=0.44). Both of those TDs are very large outliers, and in fact, when they are removed, the Pearson correlation coefficient increase to 0.51 (rejected null hypothesis) as is shown in figure 71.

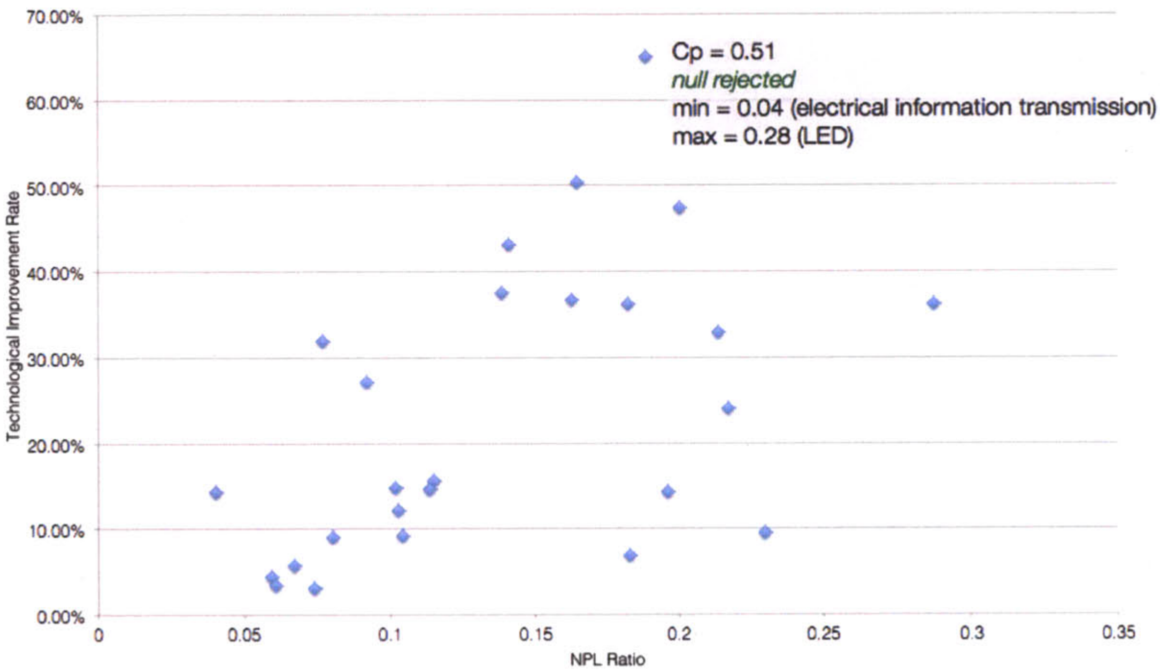



**Figure 71: TIR vs Average Non-Patent Literature Citation Ratio (2 outliers removed).**

While the relationship between TIR and NPL seems stronger with the outliers removed, the subtraction of two of the most scientifically relevant TDs does not support the reliability of this particular relationship. Despite the signal shown in figure 70, it is unlikely that the TIR and NPL ratio are strongly related. This conclusion is further supported by the fact that the other DPMs that are related to the NPL also showed no significant correlation with the TIRs as is shown in figure 72 which shows the relationship between the ratio of the number of patents in a domain with at least 1 NPL citation to the overall size of the domain.

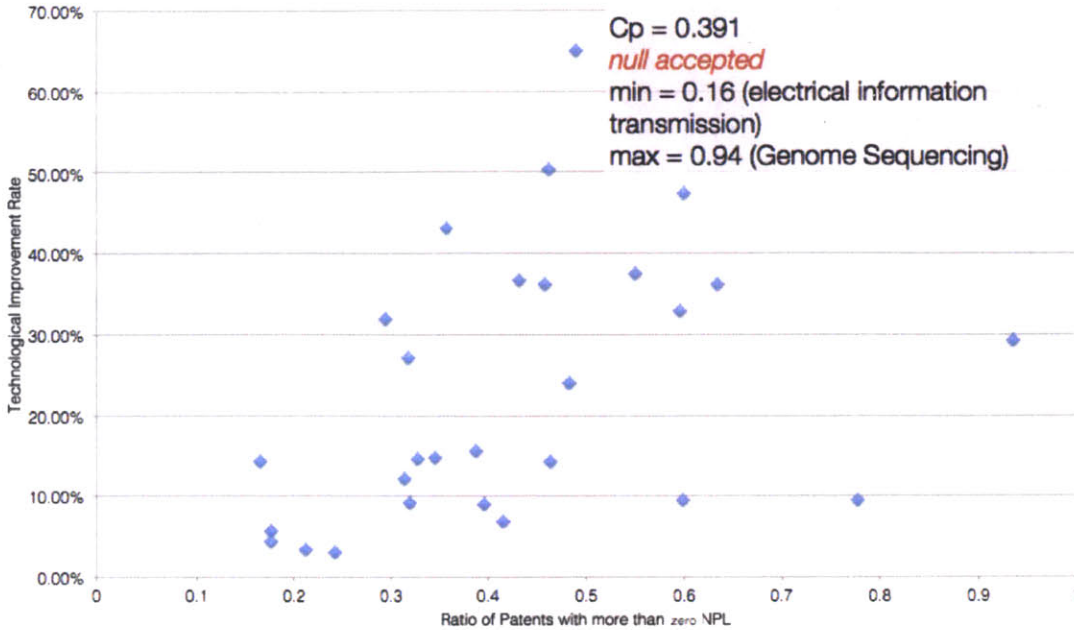



**Figure 72: TIR vs Ratio of Patents with NPL Citations > 0.**

This alternative measure of NPL also shows a very weak Pearson correlation coefficient with an accepted null hypothesis, which is consistent with the notion that NPL is not significantly correlated with technological improvement rates.

### 4.3.3.1. Top 100 NPL Ratio

As was discussed in section 4.2.5, each of the DPM is applied to the 28 patent sets and to the 28 clean top 100 most cited patents. Another way to measure the NPL citation ratio is the Top 100 average NPL citation ratio. Figure 73 shows the relationship between TIR and the average NPL ratio for the Top 100 clean most cited patents in each domain alongside the NPL citation ratio for the 28 complete patent sets.

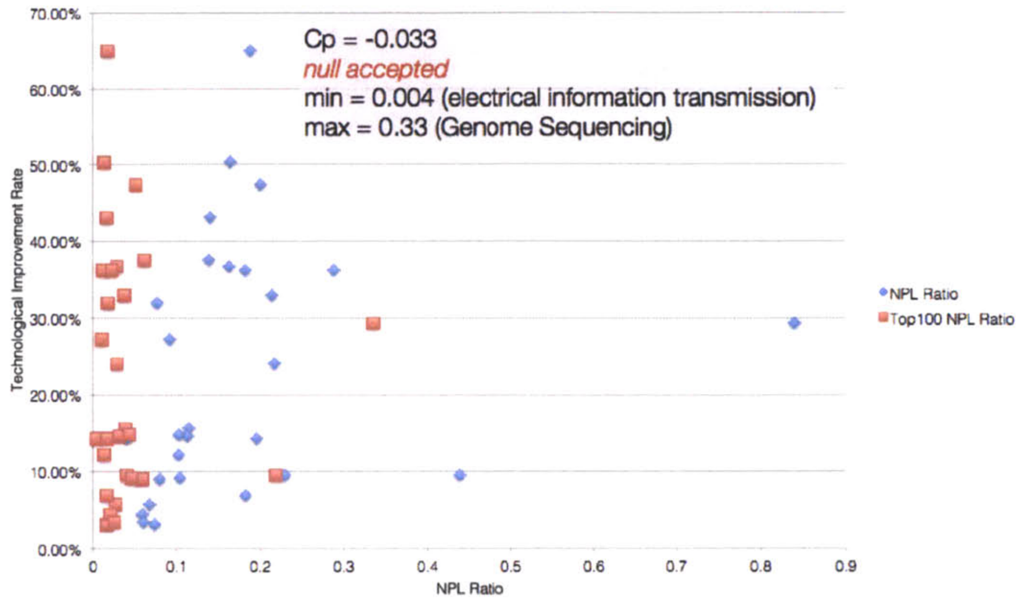



**Figure 73: TIR vs Average NPL Ratio and Top 100 Average NPL Ratio**

The Pearson correlation coefficient between the two variables is -0.033, and the null hypothesis is accepted, indicating that the very small negative correlation could be due to random scattering of the data. The combination of the statistical tests and the lack of a trend in figure 73 indicate that it is likely that there is not a relationship between the Top 100 NPL ratio in a technological domain and the associated TIR.

One of the more surprising results of this comparison is that every one of the 28 TDs showed a lower value for the Top 100 average NPL ratio than the average ratio for the entire patent set. The NPL ratios for each of the 28 TDs (both top 100 and entire patent set) are shown more clearly in figure 74. This is surprising since others (not working within defined domains as this work does) have found some correlation with forward citations and NPL whereas the top 100 patents in each domain have significantly smaller NPL (usually by more than a factor of 2). This will be discussed further in section 5.2.1.3.

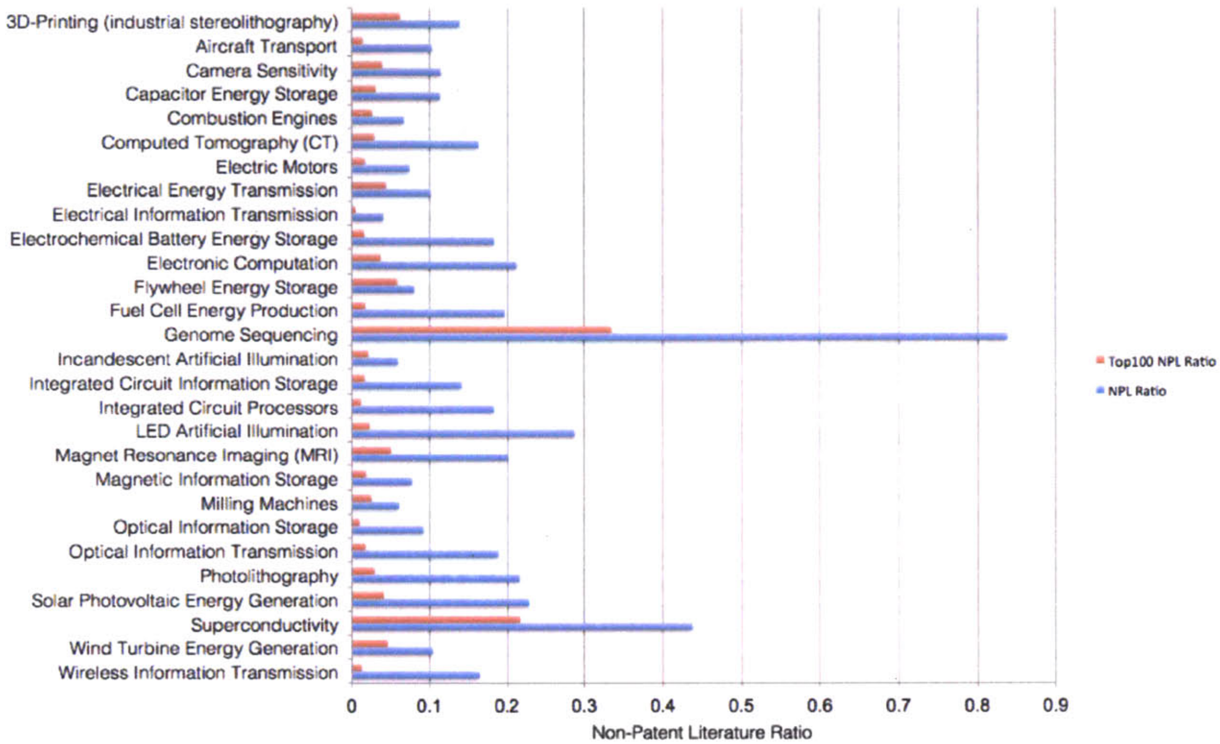



**Figure 74: Comparison of Average NPL Ratio and Top 100 Average NPL Ratio**

## 4.3.4. Hypothesis 4: Recency

Hypothesis 4 states that TDs that rely upon the most recent knowledge should improve more rapidly. The test for this is a comparison between the recency of the patents and the TIR, as is diagrammed in Figure 75.

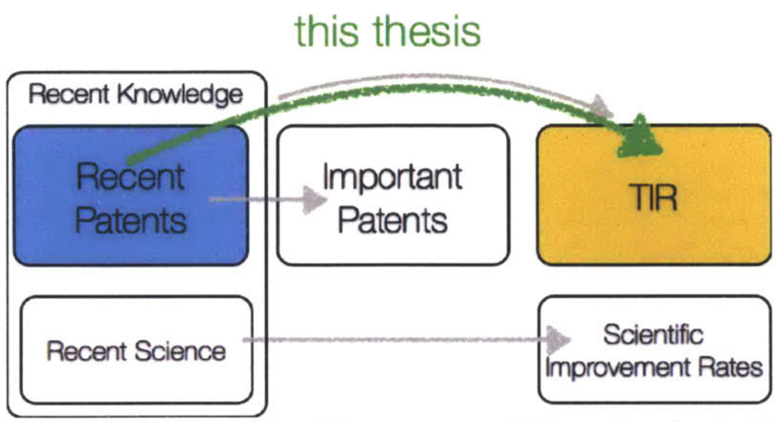


**Figure 75: Graphical Summary of Hypothesis 4 - Recency**

#### 4.3.4.1. Average Date of Publication

One of the measures used to test hypothesis 4 is the average date of publication of the patents in a TD. Figure 76 shows the relationship between TIR and the average date of publication.

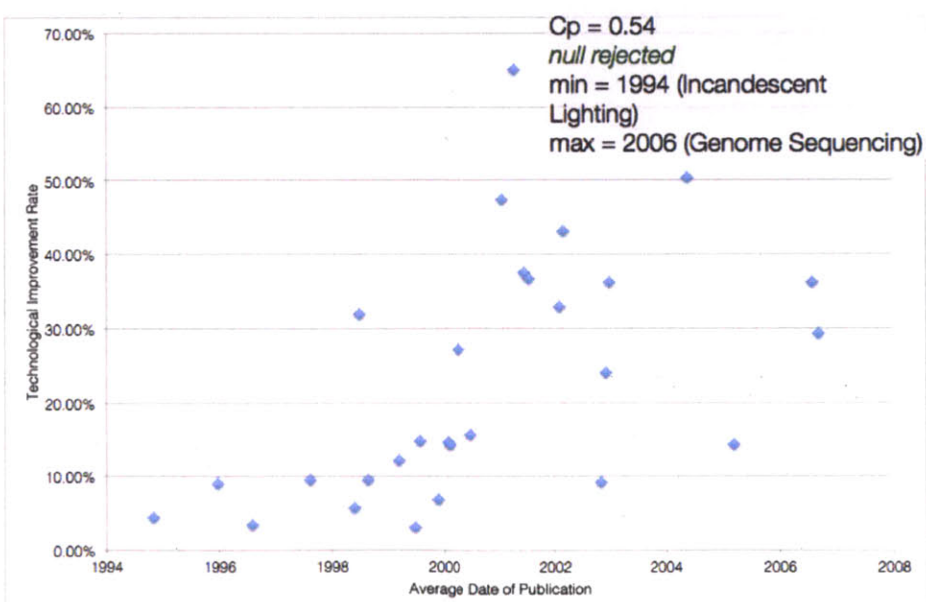


**Figure 76: TIR vs Average Date of Publication**

The Pearson correlation coefficient between the two variables is 0.54, and the null hypothesis is rejected, indicating that the correlation is unlikely due to random scattering of the data. The combination of the statistical tests and the slight trend in figure 76 indicate that there is a potential relationship between the average date of publication of the patents in a technological domain and the associated TIR.

The earliest average publication date is that of incandescent lighting and the most recent is genome sequencing, which tends to make qualitative sense when thinking about the times when those particular domains were being patented heavily. Two domains that contain many recently published patents that do not have very high improvement rates are wind turbines (TIR = 9.2% AvePub = 2002.8) and Fuel Cells (TIR = 14.4% and AvePub 2005.2). These two cases are indicative of the spread of the domains increasing over time - in general, the domains with earlier starts have a lower variation of TIRs, whereas the more recent TDs have a much wider spread of TIRs.

#### 4.3.4.2. Average Age of Backward Citations

Another measure used to test hypothesis 4 is the average age of backward citations cited by the patents in a TD as shown in figure 77.

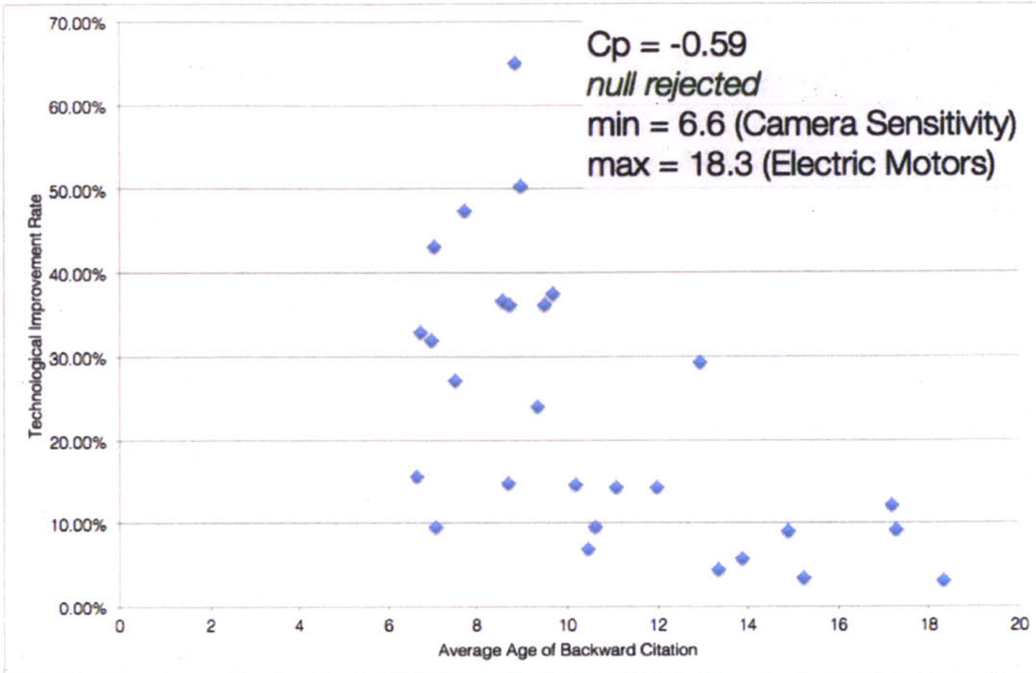



**Figure 77: TIR vs Average Age of Backward Citations**

The Pearson correlation coefficient between the two variables is -0.59, and the null hypothesis is rejected, indicating that the correlation is unlikely due to random scattering of the data. The combination of the statistical tests and the slight visible trend in figure 77 indicate that there is a potential relationship between the average age of backward citations per patent in a technological domain and the associated TIR. The correlation coefficient is strongly negative in this case, which is consistent with H4 that more recent knowledge correlates with higher TIRs. The apparent "lower limit" on age of backward citations at ~6.5 years is interesting and it would be speculation to say it simply reflects publication timing realities.

#### 4.3.4.3. 3-Year Price Index

Another measure used to test hypothesis 4 is the Price Index applied to patents as shown in Figure 78.

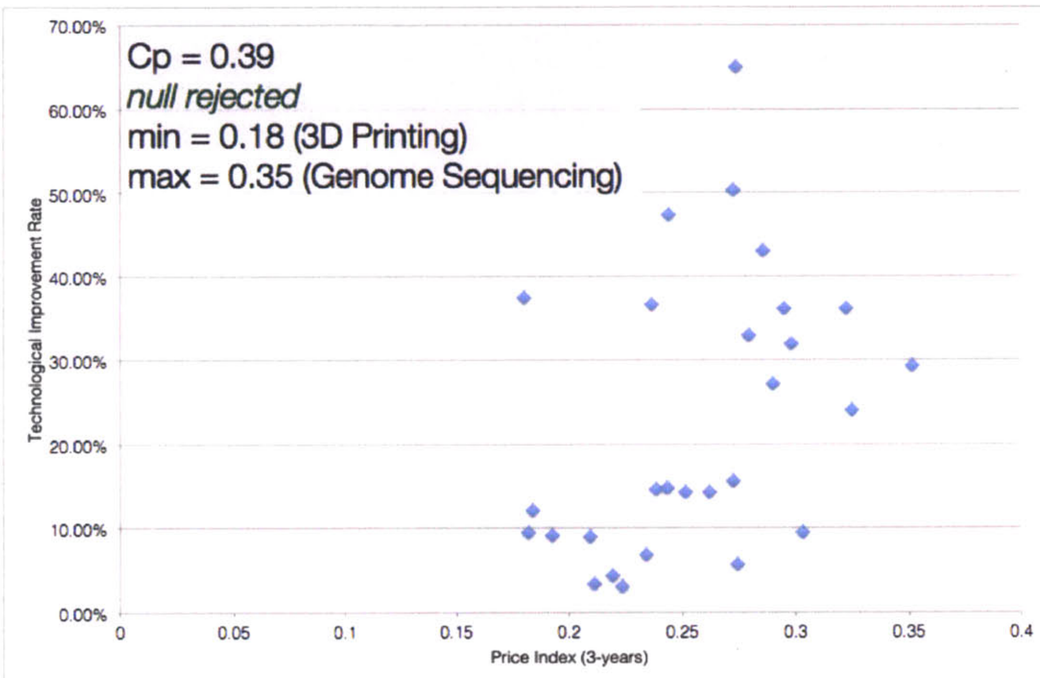



**Figure 78: TIR vs Price Index (3-Year)**

The Pearson correlation coefficient between the two variables is 0.39, and the null hypothesis is rejected, indicating that the correlation is unlikely due to random scattering of the data. The combination of the statistical tests and the slight trend in figure 78 indicate that there is a potential relationship between the average price index of the patents in a technological domain and the associated TIR.

While both are fast improving technical domains, Genome Sequencing and 3D printing represent the max and the min of the average price index per TD respectively. Also, the TD with the highest TIR (Optical Telecom) has a mid-range value of the Price Index of 0.27.

The 5-year price index was also tested and correlated nearly perfectly with the 3-year price ratio (Cp = 0.99, null hypothesis rejected), due to the extreme similarity between the metrics, only the 3-year price ratio is used throughout the thesis.

#### 4.3.4.4. Average Number of Forward Citations within 3 years of publication

A measure used to test both hypothesis 4 and hypothesis 2 is the average number of forward citations received within 3 years of publication. The results are shown in Figure 79.

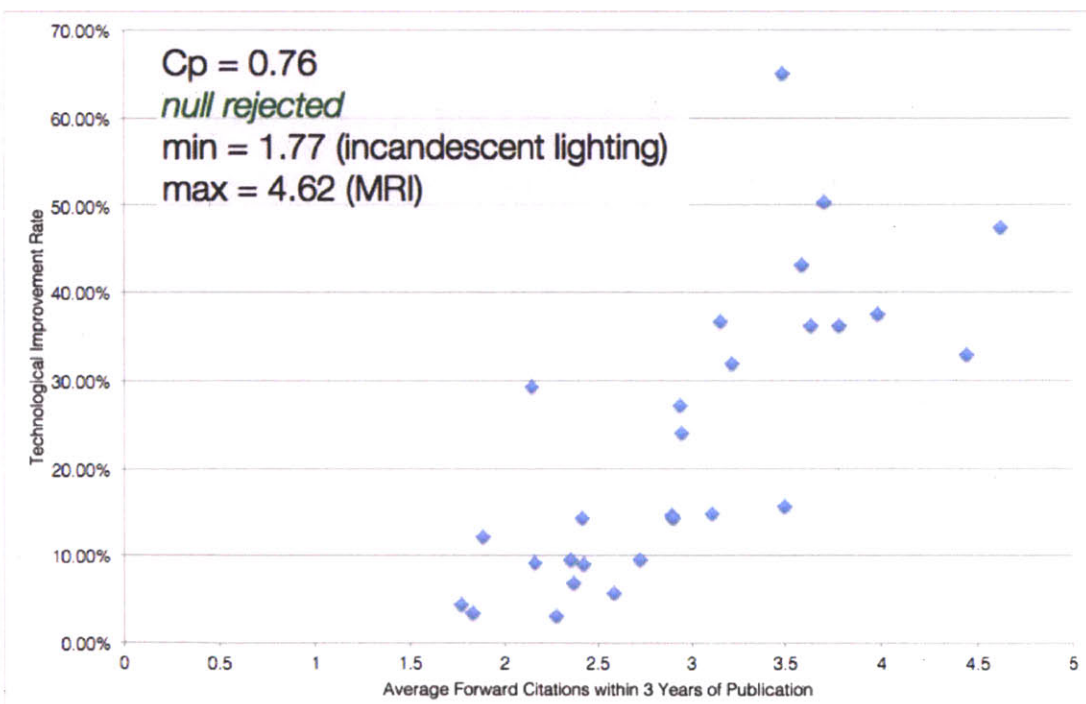



**Figure 79: TIR vs Average Number of Forward Citations Received within 3 Years of Publication**

The Pearson correlation coefficient between the two variables is 0.76, and the null hypothesis is rejected, indicating that the correlation is unlikely due to random scattering of the data. The combination of the statistical tests and the strong trend in figure 79 indicate that there is likely a strong relationship between the average number of forward citations received within 3 years of publication in a technological domain and the associated TIR.

This metric is the numerator of the 3-year Price index and is a combination of the importance hypothesis (H2) and the recency hypothesis H4), in this respect it is something of a hybrid metric. The number of citations within 3 years is the strongest correlation between the TIR and a DPM that was discovered in the entire experiment and forms the base of the prediction algorithms that will be developed later in the results section. In addition to the number of citations within 3 years after publication, the same number can be calculated for citations within 5 years of publication, and also shows a high correlation coefficient of 0.73. The high cross-correlation between the 3 and 5-year measures indicate that they are measuring similar aspects of the inventive system and thus only the 3-year domain-patent marker is included.

### 4.3.5. Hypothesis 5: Breadth of Knowledge

Hypothesis 5 is based upon the idea that TDs that rely upon the knowledge from other domains, sometimes referred to as spillover, should improve more quickly. The hypothesis states that domains that have a broader knowledge base for its patents will have a higher TIR, as diagrammed in Figure 80.

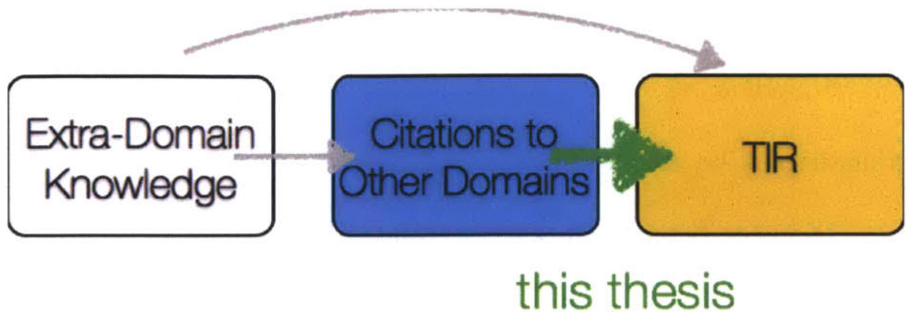



**Figure 80: Graphical Summary of Hypothesis 5 - Spillover**

### 4.3.5.1. Ratio of Backward Citations to Own Domain

One measure used to test this hypothesis is the ratio of citations to the other patents in the domain, which is the complement of the ratio of citations to *other* domains. Figure 81 shows the relationship between the TIR and the ratio of cites to the own domain.

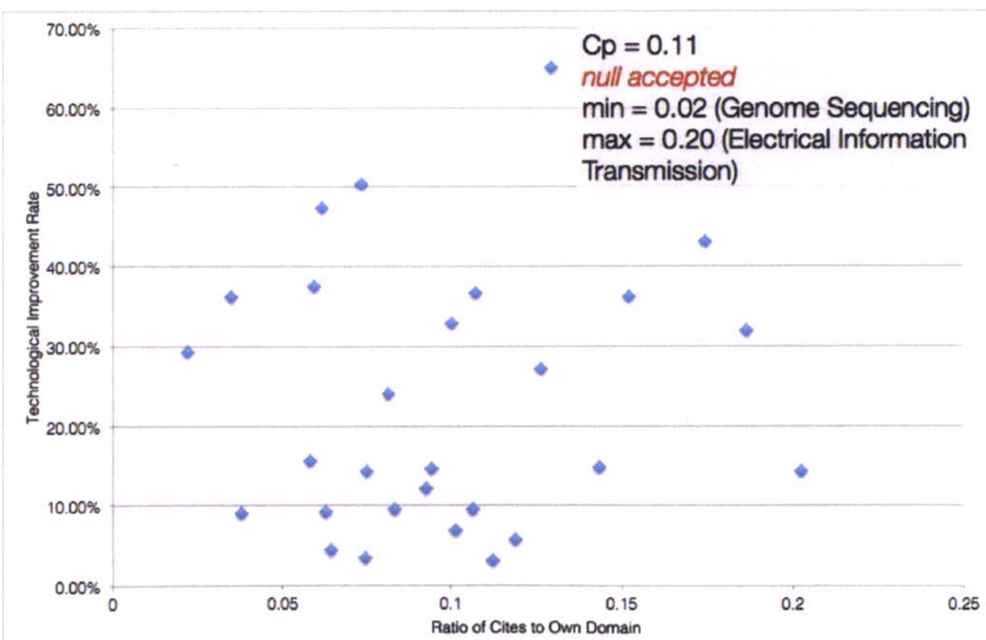



**Figure 81: TIR vs Ratio of Cites to Own Domain**

The Pearson correlation coefficient between the two variables is 0.11and the null hypothesis that the correlation could be do to the random variation in the data was accepted. The combination of the statistical tests and the lack of any discernible trend in figure 81 indicate that there is not a strong relationship between the ratio of backward citations to the own domain in a technological domain and the associated TIR.

The fact that typically only about 10% of backward citations are to patents in the domain indicates the great importance of spillover. However, this DPM is one of the weaker markers that were tested, and does not seem to show any correlation between cites to the own domain the TIR. Thus, despite the importance of spillover, it does not apparently account for differences in TIR's among domains.

#### 4.3.5.2. TIR Weighted Backward Citations

While the DPM comparing the number of cites to the own domain did not show any correlation with TIR, more elaborate methods were tested to evaluate if citing other specific TDs was correlated with the TIR. In a similar manner to the way that the overlaps of the patent sets were found, the backward citations of a particular TD were compared with the other patents sets to determine where the citations from each domain were located. This overlap ratio is calculated using equation 37, where the backward citations for the patent set of interest are denoted by *Citesi* and the 28 patent domains are represented by *Pj*.

$$\frac{Cites_i \bigcap P_j}{Cites_i} \qquad \text{(Equation 37)}$$

As an example, figure 82 shows the overlap percentages between the backward citations of the Camera Sensitivity TD with the other 27 TDs. The largest amount of overlap was 16% with IC processors, which while less than the 48.5% overlap of the patent sets as described earlier in the COM results section is still consistent in showing a strong relationship between these two domains. Only 6% of the citations are back to the camera sensitivity domain, which is the DPM described in the previous subsection.

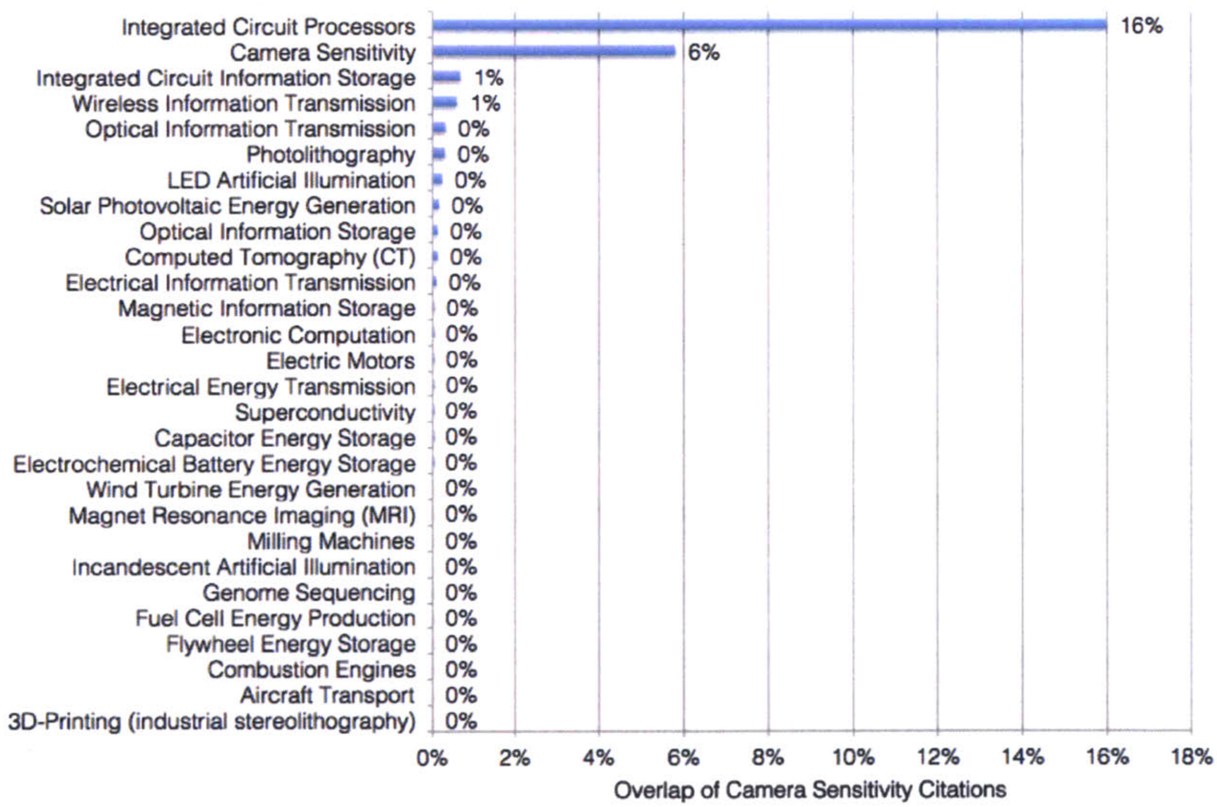



**Figure 82: Camera Sensitivity Backward Citation Overlaps**

Ultimately, very few of the patent sets are cited significantly by the camera sensitivity patents and only ~25% of the citations cite any of the 28 TD that were located for this study.

Another hypothesis was developed after the experiment had been run. Basically, it states that the TIR of a domain depends upon the TIR's of the domains cited by the domain. To test this, the values of the backward citation overlaps were multiplied by the TIR of the *TDj* as shown in equation 38 to find the individual TIR weighted patent citation overlap ratio for domain i to domain j.

$$BKWDwghtTIR_{ij} = \frac{Cites_i \bigcap P_j}{Cites_i} * TIR_j \qquad \text{(Equation 38)}$$

This value was then determined for all of the 28 domains that are classified in this study to find the total TIR weighted patent citation overlap ratio for domain *i* as shown in equation 39.

$$BKWDwghtTIR_i = \sum_{j=1}^{28} \frac{Cites_i \bigcap P_j}{Cites_i} * TIR_j \qquad \text{(Equation 39)}$$

As an example, the steps for calculating the camera sensitivity TIR weighted patent citation overlap ratio are shown in table 19.

| Domain j | Camera Sensitivity Citation Overlap | TIR for Domain j | TIR Weighted Backward Citations |
|---|---|---|---|
| 3D-Printing (industrial stereolithography) | 0.00% | 37.60% | 0.00% |
| Aircraft Transport | 0.00% | 12.20% | 0.00% |
| Electrochemical Battery Energy Storage | 0.01% | 7.00% | 0.00% |
| Camera Sensitivity | 5.80% | 15.60% | 0.90% |
| Capacitor Energy Storage | 0.01% | 14.60% | 0.00% |
| Combustion Engines | 0.00% | 5.70% | 0.00% |
| Computed Tomography (CT) | 0.12% | 36.70% | 0.04% |
| Electrical Energy Transmission | 0.02% | 14.90% | 0.00% |
| Electric Motors | 0.02% | 3.10% | 0.00% |
| Electrical Information Transmission | 0.09% | 14.30% | 0.01% |
| Electronic Computation | 0.02% | 33.00% | 0.00% |
| Flywheel Energy Storage | 0.00% | 9.00% | 0.00% |
| Fuel Cell Energy Production | 0.00% | 14.40% | 0.00% |
| Genome Sequencing | 0.00% | 29.30% | 0.00% |
| Incandescent Artificial Illumination | 0.00% | 4.50% | 0.00% |
| LED Artificial Illumination | 0.23% | 36.20% | 0.08% |
| Magnetic Information Storage | 0.03% | 31.90% | 0.01% |
| Milling Machines | 0.00% | 3.40% | 0.00% |
| Magnet Resonance Imaging (MRI) | 0.00% | 47.50% | 0.00% |

| | | | |
|---|---|---|---|
| Optical Information Storage | 0.12% | 27.10% | 0.03% |
| Optical Information Transmission | 0.32% | 65.10% | 0.21% |
| Photolithography | 0.30% | 24.00% | 0.07% |
| Integrated Circuit Information Storage | 0.67% | 43.20% | 0.29% |
| Solar Photovoltaic Energy Generation | 0.16% | 9.50% | 0.01% |
| Superconductivity | 0.01% | 9.50% | 0.00% |
| Wind Turbine Energy Generation | 0.00% | 9.20% | 0.00% |
| Integrated Circuit Processors | 15.98% | 36.30% | 5.80% |
| Wireless Information Transmission | 0.59% | 50.40% | 0.30% |
| | **24.5%** | | **7.8%** |

**Table 19: Breakdown of Camera Sensitivity TIR Weighted Overlap Citation Ratio showing how the overall value (7.8%) is calculated**

This value is calculated from equation 39 for each of the 28 TDs and is used as an additional DPM that is compared with the TIR for correlations as is shown in Figure 83.

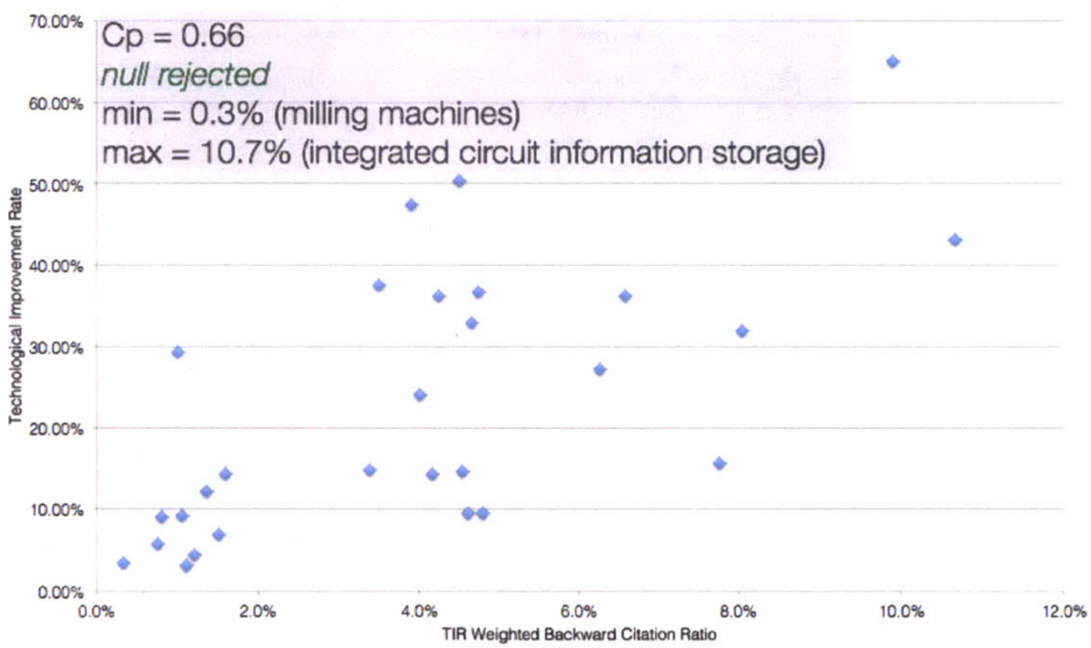



**Figure 83: TIR vs TIR Weighted Backward Citation Ratio**

This DPM has a correlation coefficient of 0.66 with the TIR and a rejected null hypothesis. The high correlation along with the visible trend indicates that higher weighted backward citations of a patent set indicates more rapid technological improvement.

It is important to note that it is an incomplete measure because it only incorporates 28 TDs out of the entire patent database. If all of the patents ever published were put into domains the sum of the overlap percentages would likely be higher than 100% (due to multiple listings) and thus the validity of this measure even with 100% coverage of patents is not certain. The research reported here neither proves nor disproves that TIR's can be calculated as a weighted sum of all their backward citations weighted by the TIR of the cited domain. This interesting possibility will be discussed further in section 5.2.1.5.

Additionally, this measure includes a bit of circular reasoning in that it includes multiplying by the self-citations of a domain by the TIR of the domain. Due to the limited coverage of our selection of patents into technological domains, a majority of the backward citations that can be binned into a technology are to the own domain (i.e. solar PV backward citations are mostly categorized as solar PV patents – while still only making up ~10% of the total number of backward citations). This suspicion is confirmed by removing the self-citation data and re-calculating Equation 39, which results in a lower Pearson correlation coefficient of 0.042 with the null hypothesis accepted. This potential for circular reasoning does not immediately eliminate the usefulness in understanding reliance on backward citations, however may be less reliable if used to predict the TIRs of other technological domains as is discussed in section 4.3.7.3.

## 4.3.6. Other DPMs

Most of the DPMs represent specific hypotheses, additional DPMs were tested as mentioned in section 3.5.6. Of the extra DPMs, the only strong positive correlation with the TIRs came from the length of title (in number of characters) as shown in figure 84.

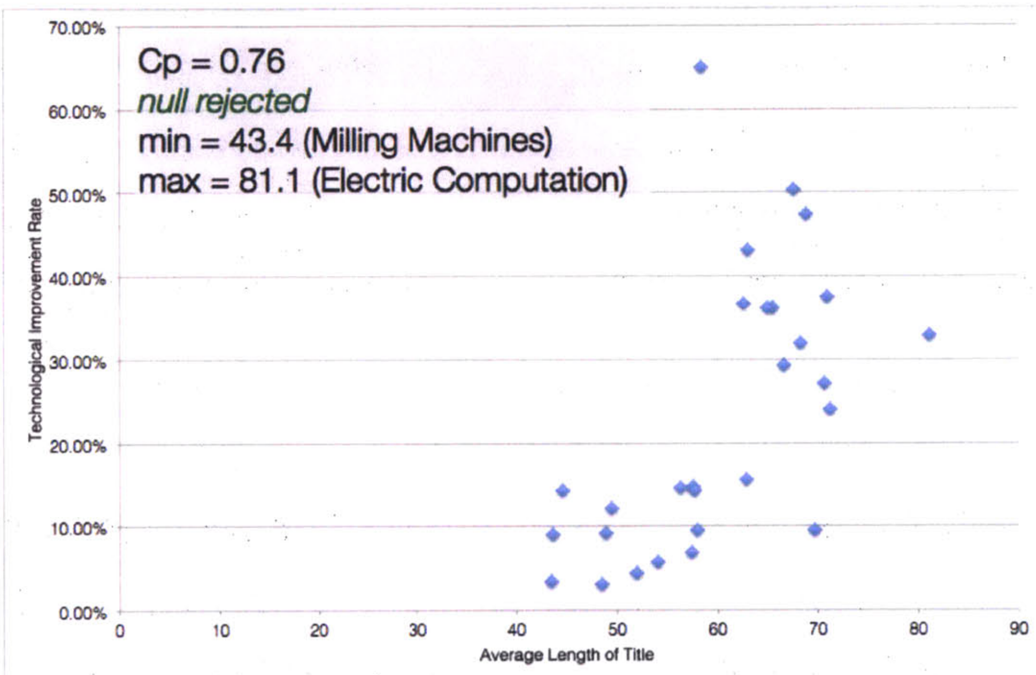



**Figure 84: Length of Title (in number of characters) vs TIR**

While the relationship between the average length of title of the patents in a domain and the TIR may seem unusual, there is some theoretical support behind this relationship from Lucio-Arias and Leydesdorff (2009). When a domain is moving more rapidly, additional context may be necessary when publishing a patent to make it's specific use more clear. In fact, this metric has a high correlation with $CIT_3$ (Cp = 0.57), which is consistent with more early citations to a paper in a field requiring more context for differentiation and clarity.

## 4.3.7. Predictive Capabilities

Over 40 different DPMs were tested for correlation against the TIRs and are listed in appendix C. Through testing of the 5 hypotheses, several strong correlations were discovered. In this section these correlations will be subjected to robustness tests based upon the domains used in the study and the length time of the study. Finally, a combination of the DPM will be combined in a number of ways to predict TIR rates using regressions.

#### 4.3.7.1. Random Selection of Half of the Domains

In order to ensure that the correlations between the DPMs and the TIRs were not simply a result of the specific set of 28 domains selected for this study, a stringent domain-based robustness test was applied to the DPMs. In order to complete this test, the set of 28 domains was randomly separated into 2 sets of 14 domains (with no TDs repeated twice) and the correlation coefficients were re-calculated using only 14 TDs each time. This trial was then completed 10 times for a total of 20 different sets of 14 TDs and corresponding correlation coefficients. An example using the TIR Weighted Backward Citation Overlaps is shown in table 20.

| | Random Trial 1 | Random Trial 2 | Random Trial 3 | Random Trial 4 | Random Trial 5 | Random Trial 6 | Random Trial 7 | Random Trial 8 | Random Trial 9 | Random Trial 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| 14 TDs | 0.704 | 0.820 | 0.596 | 0.747 | 0.658 | 0.519 | 0.471 | 0.545 | 0.692 | 0.833 |
| Other 14 TDs | 0.636 | 0.512 | 0.748 | 0.641 | 0.694 | 0.819 | 0.814 | 0.806 | 0.721 | 0.705 |

**Table 20: Random Domain Robustness Testing of the TIR Weighted Backward Citation Overlap Ratio Correlation with TIR**

The mean and standard deviation of the values were taken to determine the robustness, with the lower standard deviation values indicating higher robustness to the selection of different domains. Table 21 shows the summary of the domain robustness for the 6 DPM of most interest.

| Patent Specific Metrics | Average Correlation | Standard Deviation |
|---|---|---|
| Average Cited by within 3 years | 0.765 | **0.067** |
| Average publication year | 0.544 | **0.092** |
| Average Length (in characters) of Title | 0.604 | **0.110** |
| TIR Weighted Backward Citations | 0.684 | **0.112** |
| Average number of forward citations | 0.476 | **0.183** |
| Price Index (3 years) | 0.405 | **0.150** |
| % of patents with cited by over 20 | 0.351 | **0.245** |

**Table 21: Summary of Domain Robustness Analysis for the most highly correlated 7 Domain Patent Markers**

Given the severity of the test in removing ½ of the domains, there is quite good consistency in almost all of the effects of the metrics on the rate of improvement. In particular, the DPM of average forward citations within 3 years of publication is remarkably consistent across 20 different correlation values, indicating that the strength of that signal is not due to the selection of these specific 28 TDs. The least consistent metric was found to be the % of patents with over 20 forward citations which shows much more variation under this robustness test (it has a much higher standard deviation) and thus it is more likely that it is affected by the selection of specific data.

#### 4.3.7.2. Time Dependence of the DPMs

A second robustness test also tests the predictive capability of the correlations by testing how sensitive the DPM correlations were to variations in time. In order to do this, the DPMs were analyzed for only patents from a variety of time frames that were less than the total time frame. The time frames were analyzed to see how far back from 2013 they could be and still find similar correlations as the DPMs show during the entire time frame (1976-2013). The results of these time based tests for the two strong signals that also show high robustness as shown in table 22.

| | All Data 1976-2013 (38 years) | 2001 (26 years) | 1991 (16 years) | 1981 (6 years) |
|---|---|---|---|---|
| **Cited by Within 3 Years** | Cp = 0.76 *rejected* | 0.72 *rejected* | 0.52 *rejected* | 0.3 *accepted* |
| **Average Publication Year** | Cp = 0.54 *rejected* | 0.53 *rejected* | -0.07 *accepted* | -0.18 *accepted* |
| **Title Length** | Cp = 0.76 *rejected* | 0.42 *rejected* | 0.24 *accepted* | 0.11 *accepted* |

**Table 22: Time Robustness Analysis for 2 DPMs showing Pearson correlation and hypothesis acceptance/rejection at 95% confidence**

The three DPMs show very similar correlation values for 1976-2001 than they do for the entire data set of 1976-2013, indicating that the metrics have predictive capabilities of 12 years into the future. When the DPM were analyzed for patents before 1991, the correlation value of the cited by within 3 years DPM shows a slight drop-off (from 0.72 to 0.52) and the average publication year and average title length DPMs show a significant drop-off and has an accepted null hypothesis which renders the signal void. The decrease in reliability of this data in 1991 is at least partly due to the fact that many of the domains were very young at that point in time and thus would have very few patents to analyze. The data from the 1976-1981 timeframe shows that all three DPMs are non-signals for the short 6-years timeframe. Ultimately the two strongest and most robust DPMs are robust to time up to 12 years prior to the experiment, showing a promising amount of predictive capability.

#### 4.3.7.3. Regression Models

The strongest DPMs can be used to construct regressions to predict the TIR using only the patent data. Linear regressions were performed on the strongest DPMs in order to create predictive models. Models using other significantly correlated DPMs were constructed as well and none showed as strong of predictive power as the three shown in this section. The other models and their statistics are shown in appendix D.

Predictive Model 1 is the most basic model and uses only the average number of forward citations within 3 years as the sole predictive variable and is shown in equation 40.

$$TIR = -0.2562 + 0.1643 * FwdCit_3 \qquad \text{(Equation 40)}$$

The $R^2$ for this model is 0.58, which shows a significant amount of predictive power for using only one variable, which speaks to the strength of the $FwdCit_3$ DPM as a predictor of the TIR.

Each of the coefficients that make up the predictive equations also have p-values that give an indication of how reliable they are, the lower the p-levels are the more likely they are to be reliable in conjunction with the other coefficients, in general p-levels lower than 0.05 are better. Table 23 shows the confidence levels for each of the coefficients in model 1.

| | Value | p-level |
|---|---|---|
| Intercept | -0.2562 | 0.00253 |
| Average Cited by Within 3 Years | 0.1643 | 0 |

**Table 23: Coefficient p-levels of Predictive Model 1**

The p-values for both the intercept and the $FwdCit_3$ variable are far below the 0.05 acceptance level.

Predictive Model 2 takes into account the average number of forward citations within 3 years and adds the average publication year of the patents and is shown in equation 41.

$$TIR = -31.12 + 0.141 * FwdCit_3 + 0.015 * PubYear \qquad \text{(Equation 41)}$$

The $R^2$ for this model is 0.64, which is slightly stronger than model 1, and uses an additional variable. Table 24 shows the confidence levels for each of the coefficients in model 2.

| | Value | p-level |
|---|---|---|
| Intercept | -31.12 | 0.0468 |
| Average Cited by Within 3 Years | 0.141 | 0 |
| Average Publication Year | 0.015 | 0.0485 |

**Table 24: Coefficient p-levels of Predictive Model 2**

The p-values for both intercept and the $PubYear$ variable are slightly below the 0.05 confidence value and thus are reliable, but not as reliable as the coefficients in model 1.

Predictive Model 3 takes into account three of the most highly correlated DPMs. Predictive model 3 includes the DPMs in model 2 and adds the TIR weighted backward citation ratio, as is show in equation 42.

$$TIR = -34.60 + 0.912 * FwdCit_3 + 0.017 * PubYear + 2.11 * BwdCitOverlap_{TIRWeighted} \qquad \text{(Equation 42)}$$

The $R^2$ for this model is 0.72, which is slightly stronger than model 2, and uses three total variables. Table 25 shows the confidence levels for each of the coefficients in model 3.

| | Value | p-level |
|---|---|---|
| Intercept | -34.60 | 0.017 |
| Average Cited by Within 3 Years | 0.912 | 0.008 |
| Average Publication Year | 0.017 | 0.017 |
| TIR-weighted Backward Cites | 2.11 | 0.015 |

**Table 25: Coefficient p-levels of Predictive Model 3**

The p-values for the intercept and all three variables are much below the 0.05 confidence value and thus can be considered reliable.

Ultimately Model 3 shows the highest prediction capability with a very strong $R^2$ of 0.72. The model involves three variables that take into account importance, recency and source of knowledge. While Model 3 provides the best prediction power, it should be noted that Model 1 is astonishingly accurate for only using one DPM for its prediction. Additionally, the average cited by with 3 years DPM shows the most robustness to varying domains and timeframes, and even alone can be considered a strong indicator of technological improvement. As was mentioned in section 4.3.5.2, there may be some circular reasoning in the use of TIR-weighted backward citations because of the use of the self-citations multiplied by the TIR of the domain. When this variable is used in predicting the TIR of other domains, the values for TIR-weighted backward cites are likely to be much lower because the self-citations are not multiplied by any value because there exists no empirical evidence. It is for this reason that for predicting technical improvement rates for new domains based solely on patents that Model 2 is likely to be the most useful.

## 4.4. Various Forms of Trends in Technical Performance Increases

While most of the comparison of the domain patent markers was done to discover more about the causes of different technological improvement rates between technological domains, the creation of 28 mostly relevant and complete patent sets that represent specific TD's allowed for the testing of how a technology may improve within one TD. Specifically, while the number of patents within a TD did not correlate highly with the varying TIRs, one might be able to replace time as a dependent variable and use patent counts when tracking the performance of a technology over time. This is similar to the extensive literature that uses production as a replacement for time in tracking progress as is discussed in section 2.1 of this thesis. In this section, several different types of effort variables for tracking the performance of Integrated Circuit Processors will be explored and compared.

### 4.4.1. Demand

This subsection will explore the relationship between the number of units produced in the semiconductor industry with time and technology performance.

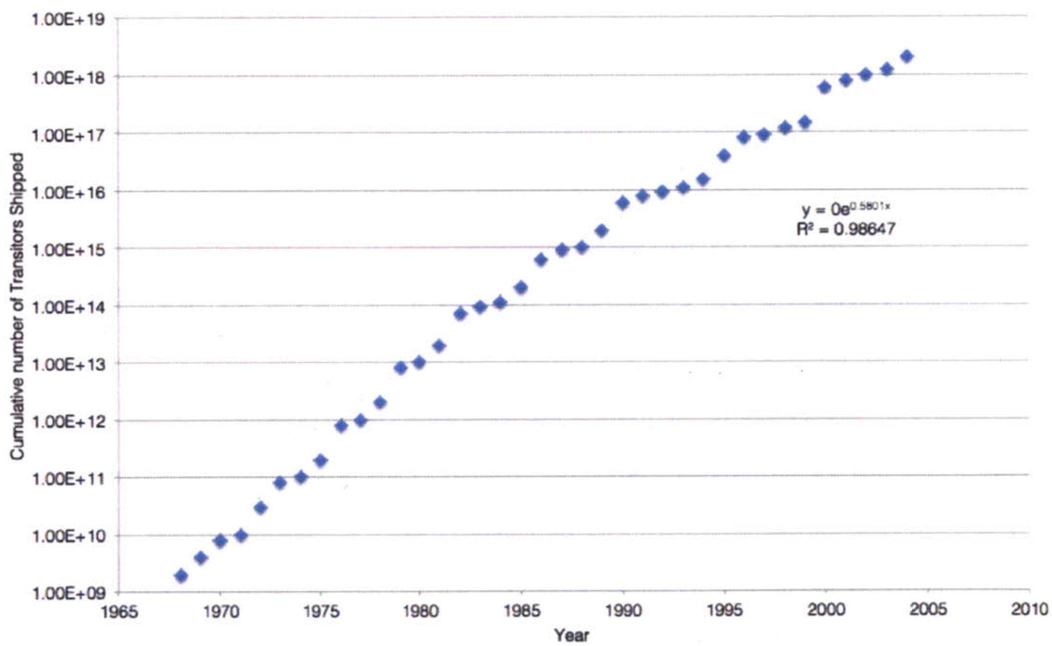



**Figure 85: Log IC Production vs Time - adapted from (Moore, 2006)**

The $R^2$ of the exponential fit is a very high 0.986 and the increase rate of the production is 58.01% per year, which is much larger than the revenue or patent growth rates.

The FPM are plotted on a log-log plot (after Wright) against cumulative industry shipments per year and is shown in figure 86.

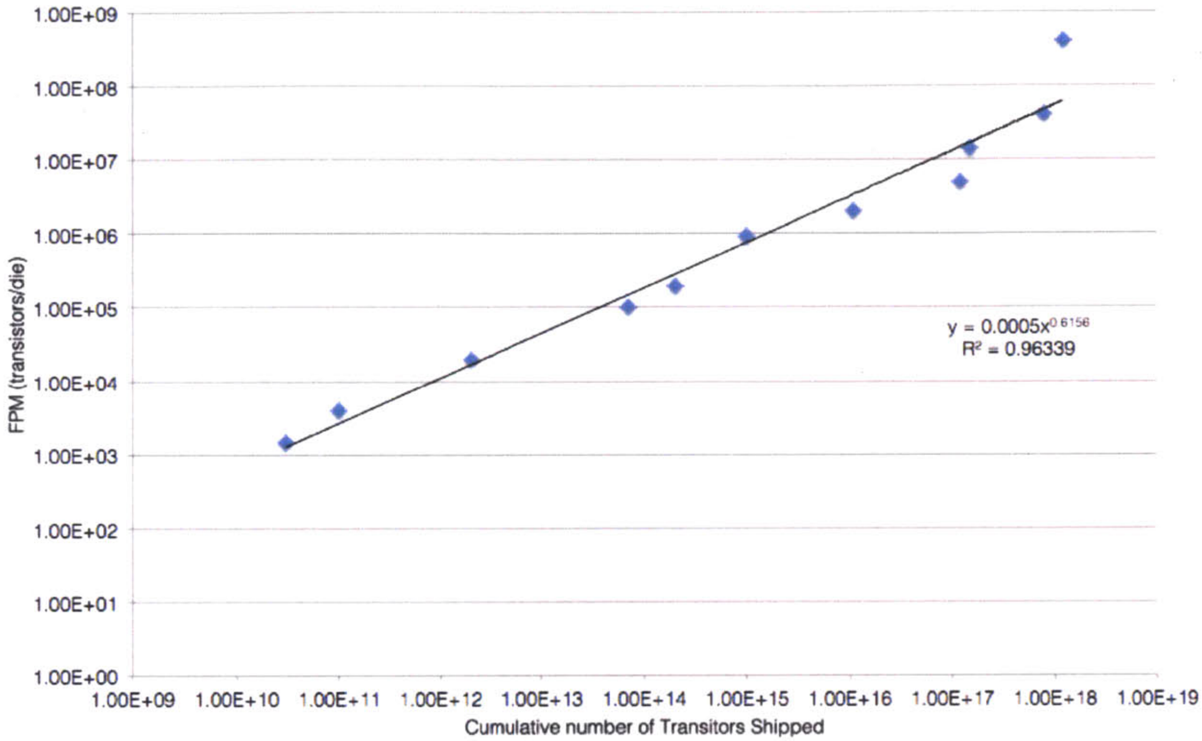



**Figure 86: Log transistors per die vs Log Cumulative Industry Production**

In a similar relationship to that of the two previous effort variable, the industry production vs FPM also exhibits a power law relationship with an exponent of 0.62 and an $R^2$ of 0.96. The exponent of the power law is the same as the variable x in equation 2 (Wright's law).

## 4.4.2. Revenue

Another effort variable that can be used to track the performance of a technology is the revenue of a particular industry, which is often seen a proxy for R&D spending. The increase in revenue of the IC industry is shown in figure 87.

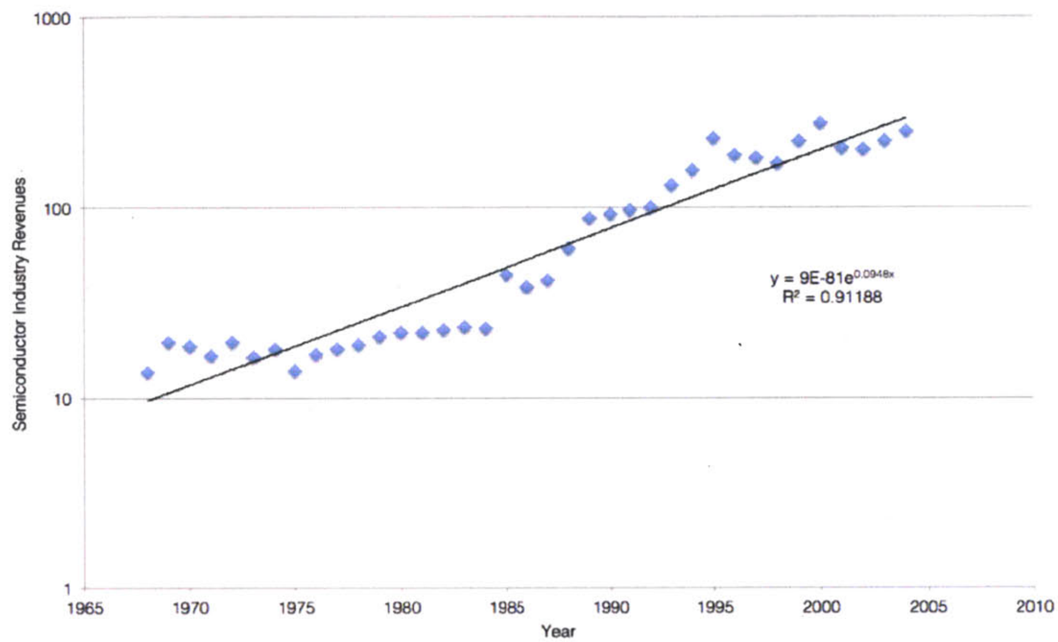



**Figure 87: Log Cumulative IC Revenues vs Time - adapted from (Moore, 2006)**

The $R^2$ of the exponential fit is 0.912 and the rate of the revenue increase is 9.48% per year.

Again following Wright (1936), the FPM can be plotted on a log-log plot against industry revenue per year as is shown in figure 88.

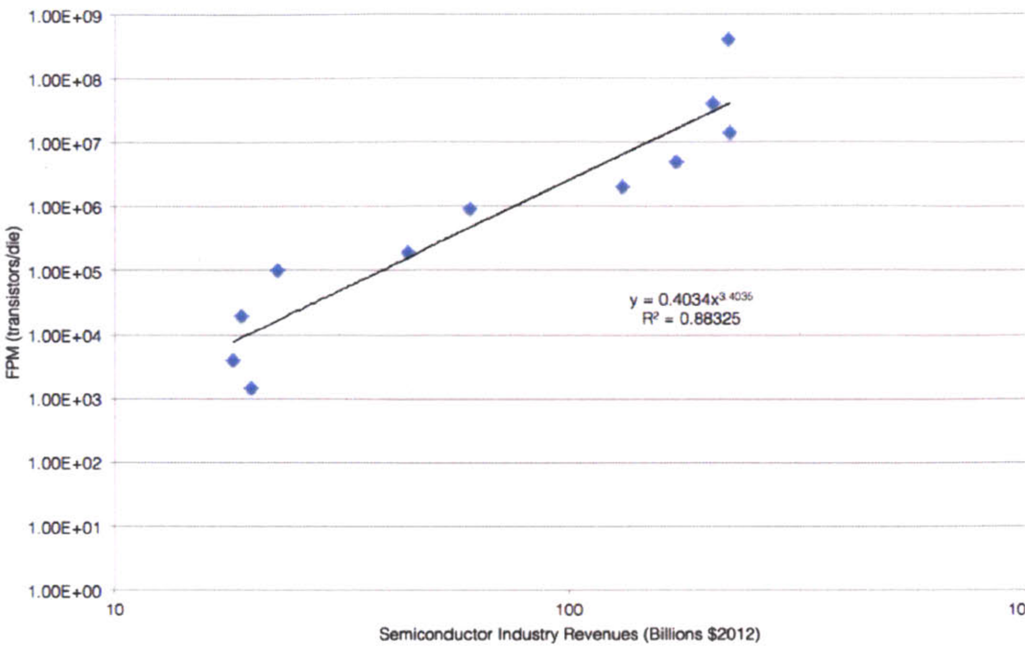



**Figure 88: Log Transistors per die vs Log Industry Revenues**

In a similar relationship to that of the cumulative patents vs FPM, the industry revenues vs FPM also exhibits a power law relationship with an exponent of 3.4 and an $R^2$ of 0.88.

### 4.4.3. Patents

While the DPM of simple patent count treated the number of patents in a domain as a simple number, in all cases the number of patents per year varied with time. In a number of cases, the number of annual patents in a technological domain increases exponentially over time. Figure 89 shows the cumulative number of patents published in the IC processor domain since

1980. The first few years (1976-1979) are not included as they are unreliable because 1976 was the first year of the data set.

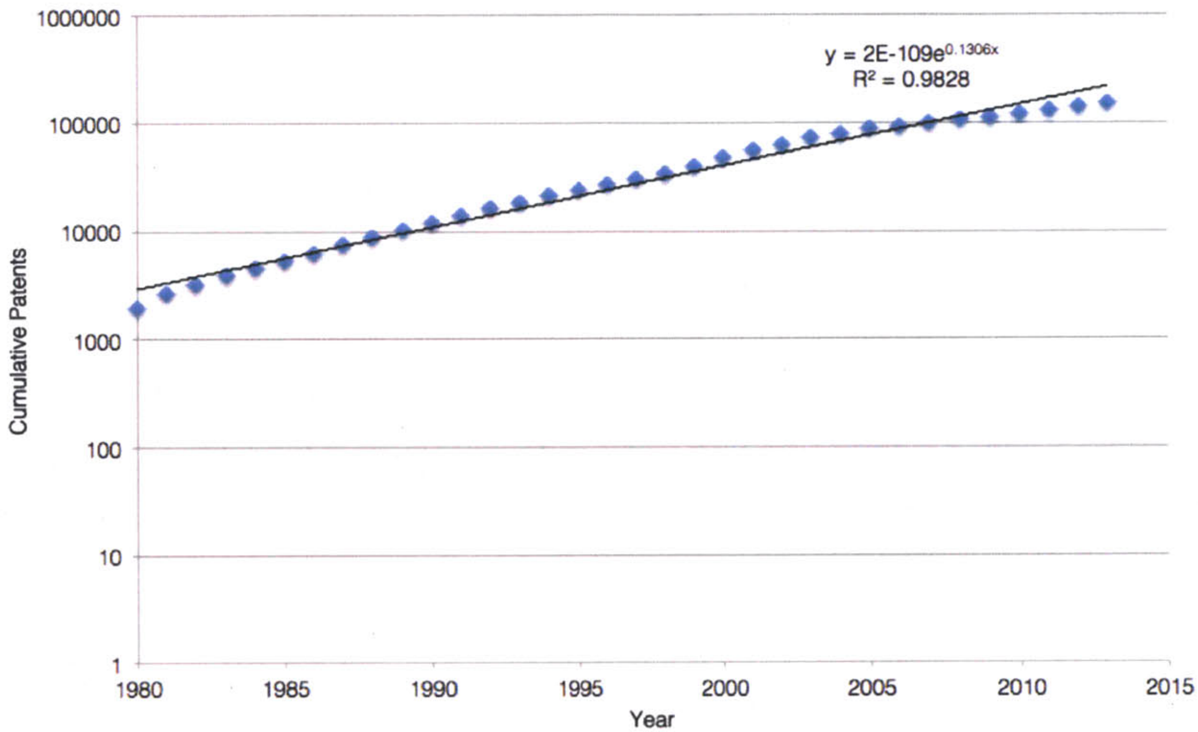



**Figure 89: Log Cumulative IC Processor Patents vs Time (patent data starting in 1980, the first data point includes patents issued from 1976-1980)**

The $R^2$ of the exponential fit is very high at 0.98 and the cumulative patents increase approximately 13% per year.

Following Wright and the extensive work using cumulative production as the independent variable, the FPM is plotted on a log-log plot against the number of cumulative patents per year can as is shown in figure 90.

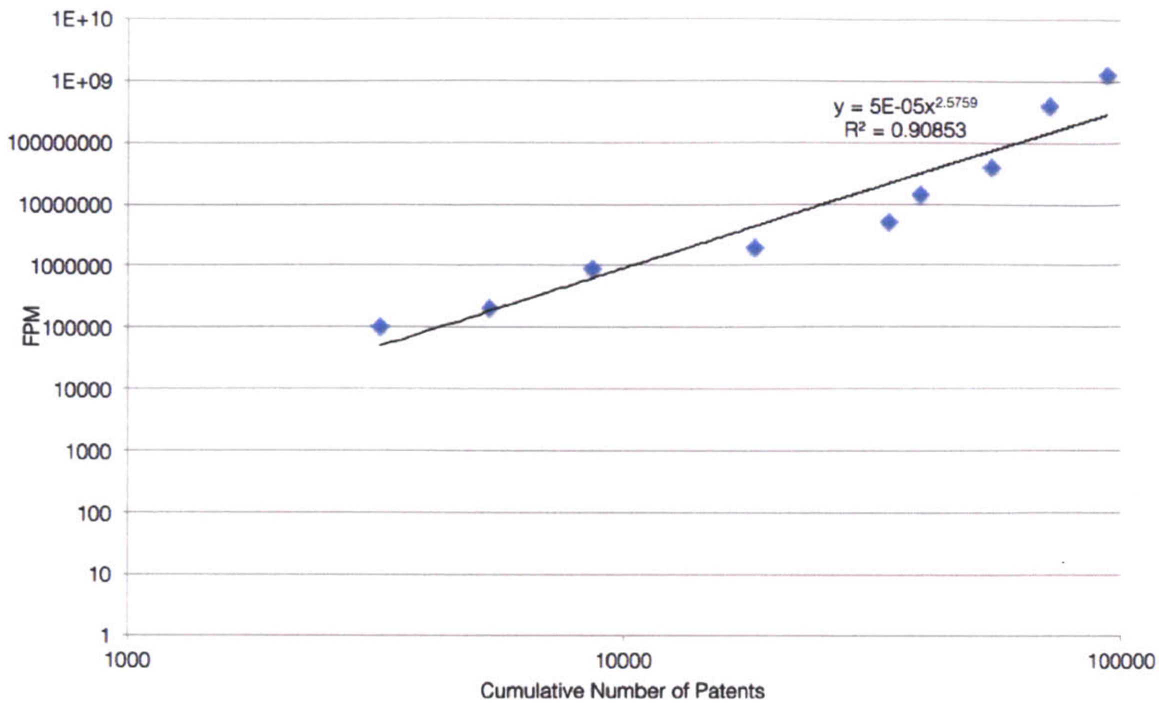



**Figure 90: Log Transistors per die vs Log Cumulative IC Processor Patents**

Because the increase in cumulative patents vs time is exponential and the increase of FPM vs time is also exponential, the relationship between the FPM of IC Processors and the cumulative patents is a power law relationship (Sahal, 1979). The $R^2$ of the power law regression is 0.91 with a power law coefficient of 2.57.

## 4.4.4. Sahal's Relationship

Each of the effort variables shows a different improvement rate over time and a different power law exponent when compared with the FPM of integrated circuits. As is discussed in more depth in section 2.1 of this thesis, Sahal (1979) showed that these effort variables can be equated by multiplying the exponential increase rate over time (Beta) by the power law exponent (alpha) with the resulting value equaling the TIR of the technology as shown in equation 43.

$$TIR = alpha * beta \qquad \text{(Equation 43)}$$

This relationship can be tested by multiplying the alpha (exponential improvement over time) and the beta (power law exponent with FPM) values that were collected for each of the three effort variables. Table 26 shows the results of the empirical tests of Sahal's relationship using patents, revenue, and production as the effort variables.

| | alpha | beta | alpha*beta |
|---|---|---|---|
| Patents | 2.57 | 0.13 | **0.33** |
| Revenue | 3.4 | 0.095 | **0.32** |
| Production/Demand | 0.6 | 0.59 | **0.35** |

**Table 26: Empirical test of Sahal's Relationship; Alpha x beta should equal k which empirically is .36 for this domain-metric pair in good agreement with the estimates from Sahal.**

The values of alpha*beta are all very near the TIR of IC Processors (0.36) despite the fact that the respective alpha and beta variables vary considerably. This empirical test provides empirical support for Sahal's relationship and lends further credibility to the theory that technological performance can be tracked against a number of different variables which are equal when compared to each other as was strongly stated by Nagy et al (2013).

### 4.4.5 Alpha and Beta for all 28 Domains

While the patent increase rate (Beta) and the power laws relationship between patents and performance (alpha) fit well for the integrated circuit processors domain, this was not universally the case. Figure 91 shows the R2 values for the Alpha, Beta and TIR for each of the 28 TDs.

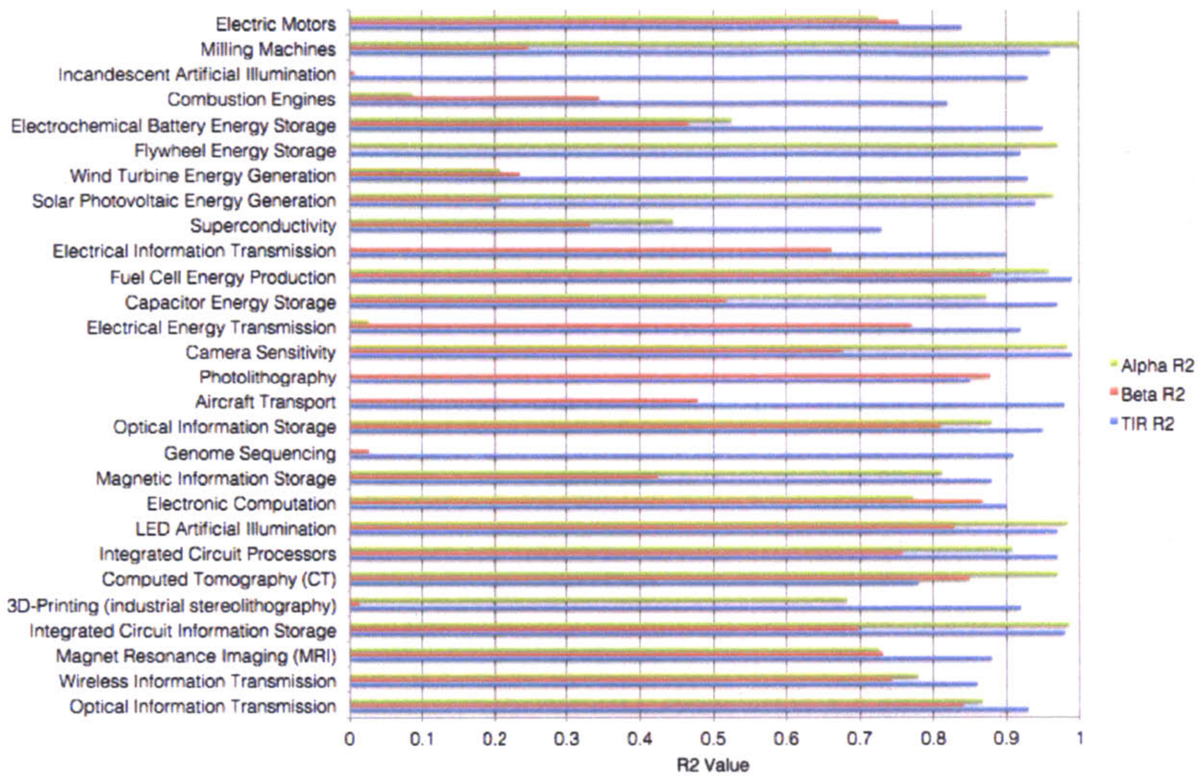



**Figure 91: R2 for TIR, Alpha and Beta for Patents for all 28 TDs**

The TIR is consistently the most reliable measurement of technological improvement with an average R2 of 0.9125. The Alpha power-law relationship is often similar in reliability to the TIR, however is dragged down by five cases that could not be computed (and thus have R2 of 0) due to declining patent increase rates leaving the mean R2 value of alpha at 0.612. The exponential increase of patents over time has a mean R2 of 0.54.

Thus, while the patent based measure of effort for tracking technological change is possible and is accurate for some domains, it is far less broadly applicable over a wide variety of technical domains.

## 4.5. Summary of Results

The approach of this research is to analyze technological improvement rates over many different technological domains and compare those values with markers derived from patents. The results of the study provided 28 patents sets and corresponding technological improvement rates for a wide variety of technologies, insight into why those different TDs may improve at different rates, an algorithm to estimate TIRs based upon patent analysis and empirical support for the theory that several different measures of effort are nearly transformations of each other based upon Sahal's relationship. This final sub-section of the results will provide a short summary of all of the major findings of this research.

First, the TIRs for each of the 28 domains was subjected to a number of reliability filters to determine the most complete and reliable TIRs for each TD as are shown in figure 58 in section 4.1.2.2. Next, the classification overlap method was used to select 28 patent sets to represent each TD that were relatively complete and relevant as is shown in figure 61 in section 4.2.6. The domain patent markers were then correlated with the TIRs in order to better understand the variation of TIRs between TDs.

| | Correlation with TIR | Null Hypothesis |
|---|---|---|
| **Hypothesis 1: Effort** | | |
| Simple Patent Count | 0.33 | accepted |
| **Hypothesis 2: Breakthroughs** | | |
| Average Number of Forward Citations | 0.48 | rejected |
| % of Patents with More than 20 Forward Citations | 0.39 | |
| **Hypothesis 3: Science** | | |
| NPL Ratio | 0.2 | accepted |
| **Hypothesis 4: Recency** | | |
| Average Date of Publication | 0.54 | rejected |
| Average Age of Backward Citations | -0.59 | rejected |

| | | |
|---|---|---|
| Price Index (3-year) | 0.39 | rejected |
| Average Number of Forward Citations within 3 Years of Publication | 0.76 | rejected |
| **Hypothesis 5: Spillover** | | |
| % of Backward Citations to Own Domain | 0.11 | accepted |
| TIR Weighted Backward Citation Ratio | 0.66 | rejected |

**Table 27: Hypothesis derived domain patent markers Correlations with technological improvement rates along with null hypothesis acceptance or rejection**

The DPMs with the strongest correlation values were tested for robustness against varying domains and timeframes and then were combined into a model that can be used to estimate TIR values for a TD based upon the patent data. Three models were created and showed promising consistency with the observed TIRs and Model 2 is likely to be the most reliable for prediction purposes.

$$TIR = -0.2562 + 0.1643 * FwdCit_3 \quad (R^2 = 0.58)$$

$$TIR = -31.12 + 0.141 * FwdCit_3 + 0.015 * PubYear \quad (R^2 = 0.64)$$

$$TIR = -34.60 + 0.912 * FwdCit_3 + 0.017 * PubYear + 2.11 * BwdCitOverlap_{TIRWeighted} \quad (R^2 = 0.72)$$

Finally, the patent data was used to support the concept that technological performance can be measured against a number of different effort variables in an equal manner. The IC processors domain was used to compare 3 different effort variables, which were related using Sahal's relationship and shown to be nearly equal to the TIR of IC Processors which is 0.36 as was shown in table 26 in section 4.4.4.

# Chapter 5: Discussion and Contributions

This section of the thesis will discuss the results of the inter-domain patent analysis as compared with the technological improvement rates of the 28 domains. The results of the study will be discussed with respect to their context within overall theories of technological change and the contributions that this research has provided to the field. Finally, several practical implications of this research will be described.

## 5.1. Contributions of the thesis to the Study of Technological Change

As was foreshadowed in section 1.3 with Popper's thoughts on falsifiability, the main result of this research is not any single domain patent marker correlation with the TIR or even the predictive model that allows for the estimation of TIRs. The structure of the experiment and its implications to future technological research are possibly the most useful components of this study. This thesis is the first experiment to explain variation in TIRs for multiple domains using characteristics of the domains derived from patent data. Several statistically significant results were gathered from this research and can be used to objectively discuss theories of technological change, something that has rarely been done before in such depth. Furthermore, the 28 domains studied provide extensive breadth and should allow for an excellent base of knowledge for further objective development of technological change theories. However, to achieve this, such theories have to have a quantitative predictive framework which most do not.

### 5.1.1. Repeatability and Objectivity of performance improvement and patent sets in domains

In order to compare the TIRs and the DPMs, care was needed in order to ensure that the results were objective and repeatable. Many of the techniques used in this research were created specifically to ensure robustness of the final results.

While there have been many studies of TIRs in the past, few have tested the reliability of the TIRs as fully as in this thesis. While the $R^2$ value is a great indicator of how well the data fits a particular regression, it falls short on capturing the true reliability of a particular exponential growth rate of a technology. This study incorporated the confidence interval standard deviation as well as the point removal method to test for confidence of the fit and robustness to the removal or addition of certain points. The PRM is a statistical test that is designed specifically to be used on TIRs as there is never 100% completeness of the TIR data and therefore the robustness to added or removed points should be evaluated whenever measuring a TIR. A great amount of effort was put into testing the reliability of the TIRs in this study, and the reliability of TIRs should be emphasized more than it has been previously, potentially using the methods described in this thesis.

Locating and testing a set of patents for 28 greatly varied technological domains is another main contribution of this thesis. The COM was designed to be repeatable and robust to user input so as it can be used in a large number of different scenarios, even ones beyond the fields of technological change. In previous studies of technological change and patent analysis many of the tests were completed on a small number of highly relevant patents within one domain, or patents felt to represent a domain but without consistent relevance testing or a large number of mostly anonymous patents across many unidentified technological fields. This thesis applied those patents metrics to a large, varied set of specifically identified sets of patents as well as clean sets of the top 100 most cited patents in each TD. The surprisingly broad applicability of the COM and its usefulness in this study should not be understated, as it could provide a basis of selecting patents for study within and outside of the technological change field.

### 5.1.2. Combining Quantitative and Qualitative Results

The results of the experiment are intentionally and largely quantitative in nature so that they can provide objective measures necessary for prediction of technological change. This quantitative information is supported by large amounts of qualitative information about each TD and each DPM that can also be used in interpretations of the results of the experiment.

The significant attention paid to robustness of the TIRs, relevance of patent sets and the robustness of the experimental results from comparison of the TIR's and patents are the foundation upon which the objective findings are built. Five hypotheses derived from prevailing theories of technological change were used to arrive at reliable quantitative understanding of the cross-domain patent TIR comparisons. Due to the quantitative nature, it was possible to show whether the results were significant or potentially due to data sampling issues. The ability to test multiple theories of technological change at a statistically significant level across 28 different TDs is a significant contribution to the study of technological change. However, many of the existing theories of technological change are stated in narrative and other non-falsifiable forms. Thus, failing these quantitative tests do not entirely eliminate the theory but do indicate a need for further objective development. Further studies building from this thesis research should be able to test more theories across more domains resulting in even higher reliability quantitative data and hopefully resulting in reliable falsifiable theories of technological change.

Quantitative data provides the objective information to evaluate theories of technological change, it is important that it is used in conjunction with qualitative information available to not only test the results, but to understand them. Each of the TD patent sets contains a wealth of qualitative information contained in the text of the patents. In particular the clean sets of top 100 most cited patents in each technological domain provide a more concise means for understanding what the most important inventions were at different times that contributed to the development of the technology. The list of mostly highly overlapped patent sets provides quantitative grounding for the qualitative concept of how closely related certain TDs are to one another.

The DPMs were mostly used for the correlations between the patent sets and the TIRs; however, the specific values of some of the DPMs for each TD give a qualitative sense for how specific TDs developed. For example, it was interesting to note that the Genome Sequencing TD relied almost entirely on scientific literature for its citations, something that aligns with our intuition about the field. Another example is that incandescent lighting has the oldest average date of publication, a fact that aligns well with the US government's attempts at discouraging incandescent light bulbs over the last few decades.

This methodology places significant emphasis on objective and quantitative methods to test theories of technological change, however it also provides a significant amount of complementary subjective and qualitative evidence that are useful in conjunction with the main findings of the thesis.

## 5.2. Contributions to Technological Change Theory

One of the main contributions of this thesis is the quantitative empirical data that can be used to support (or not support) theories of technological change. Five main hypotheses were tested using a number of different DPMs across the 28 TDs. In this section, the results of these tests are individually synthesized and interpreted within the context of theories of technological change. The second sub-section will describe an analogy-based concept that is consistent with the main findings of this thesis called the 'Rising Sea Metaphor'.

### 5.2.1. Direct Implication of the Results

It is important to begin this sub-section with a clear distinction between correlation and causation. Indeed, many of the results of this research are based upon the correlation of the DPMs and TIRs of the 28 different domains, yet at this point in time there is insufficient evidence to purport that the DPMs can be causally linked to the improvements of technological domains. Even if casual links were potentially derived from a theoretical basis, in many cases the direction of causality can be questioned as well (i.e. Do highly cited new patents cause an increase in technological performance or do TDs that improve rapidly cause patents to cite more recent patents highly?). Despite the lack of casual relationship between the DPMs and the TIRs, the relationship is still useful for understanding technical change and for estimating TIRs using the predictive model.

#### 5.2.1.1. Effort

The number of patents in a technological domain shows almost no correlation with the TIRs. This very weak signal may be partially due to the fact that the patent sets are not 100% complete, however considering the large number of domains evaluated in this thesis, it is unlikely that even with perfectly complete patent sets (which would be impossible to select), that the number of patents in a TD would correlate strongly with the TIRs. This result indicates that the number of patents is a poor indicator of inventive impact and is consistent with a number of previous studies, that citations rather the number of patents in a particular TD is a much better indicator of the economic impact of inventive effort.

The patent based effort metrics for the IC Processors domain proved to be consistent with Sahal's relationship between the cumulative number of patents published in the TD and the FPM. Likewise, the other two effort variables (revenue and production) also followed Sahal's relationship as introduced in section 4.4.4. While these 3 effort variables proved to be nearly equivalent to the time based TIR for IC Processors, many of the other domains did not have reliable cumulative patent improvement rates and the correlation of FPM's with cumulative effort was not very reliable in almost ½ of the TD's which supports the theories that the various effort variables are nearly equal as mentioned in section 2.1.4. Not only does this make it not possible to test Sahal's relationship, but effort based analysis is not capable of assessing technological progress in these TD's. Importantly, when constructing the effort-based technological improvement curves, time was used as an intermediary variable (FPM vs time and patents vs time) and all of the TIR results based upon direct correlation of FPM and time were reliable enough to assess all 28 domains.

Qualitatively, it appears that Moore's law is more fundamental than experience curves for technological progress in the 28 domains. Since progress as we measure it is due to new artifacts (new designs/inventions), this result has logical appeal. In particular, it is important to note that a very small portion of the backward citations from each TD were to other patents in that particular TD. Thus, ongoing development in a particular TD is influenced significantly by improvements in other technologies so it is not logical that the specific number of patents or revenue or demand in one individual TD can sufficiently explain its performance improvement. It is for these reasons that it is likely that the most stable method of measuring the performance of a technology (where progress involves new designs/inventions) is to use time as the dependent variable. Nonetheless, in cases where the effort variables improve exponentially experience curves provide different yet equal methods of assessing technical performance improvement.

#### 5.2.1.2. Importance: Citation Frequency and networks of patents

There are several DPMs that are related to citation frequency that correlate highly with TIRs. Thus the ideas that citation count correlates with importance and that TDs with more important inventions improve more rapidly are supported by this research. The exact mechanisms in which the more important inventions influence the improvement rate are less clear. The results of this study show that the average number of forward citations correlate more strongly with TIRs that the percentage of patents with more than 20 citations as was introduced in section 2.2. This could indicate that domains are less propelled by a set of *very* important inventions and it is more likely that TDs in which the patents tend to be used more often improve more rapidly.

One of the main issues with measuring importance that is reflected in the difference between *very* important patents and a set of more important inventions on average is the unit of analysis of importance. In this experiment, patents are used as proxies for inventions and forward citations are used as a heuristic for impact; both assumptions are based on prior literature discussed in section 2.3. Many of the other theories of technological change listed in section 2.2.7 attempt to assign qualitative labels to inventions such as *breakthrough* or *radical* and it is sometimes unclear or very subjective to what should be labeled as *breakthrough* or *radical.* Many of the prior examples of these qualitatively selected very important inventions are represented by a large set of patents in this thesis, which could help explain the slight variation between the average forward citation and the percentage of patents with more than 20 forward citations. Together, these inter-related or networked sets of patents are likely capable of causing significant disruptions or forming entirely new architectures for artifacts within a TD. While these groups of inventions are often labeled as coming from a single innovation after the fact, it is often the case that they are the accumulation of a great number of inventions to create the, radical, breakthrough, "new paradigm" or "punctuated equilibrium" innovation. In fact, it is such networks of patents that cause ongoing rapid improvement suggesting that the many terms used are not in fact time bound but occur throughout time in a TD and often leverage key improvements from other domains.

The issue of correlation vs causation is especially salient when considering our strongest DPM signal- the number of forward citations within three years of publication- as it is very difficult to distinguish between the two possible causal directions: one where the inventions in a TD are used quickly and significantly after they are invented and thus the domain improves rapidly or where a domain is improving so rapidly that the information that is important is only valid for a few years after publication. This lack of casual proof applies to all of the importance related DPMs. Despite this and following the precedent set by Solow and shown in section 2.23, the correlation between highly cited inventions and TIRs is statistically significant across three importance related DPMs, providing a strong indication that TDs that improve more quickly have inventions that are used more often and particularly that the more recent inventions are used more often. Given that most of the citations from a domain are to other TDs, it is important to note that the inventions within a domain need not be used by inventions in that domain for the TD to be improving quickly suggesting that the signal is not causation. This fact lends credence to the idea that domains that are improving more quickly lead to patents that are cited more frequently by other domains than patents in a slowly improving domain.

#### 5.2.1.3. Science: NPL as an indicator

The NPL citation ratio is the main heuristic used to represent scientific reliance of a patent as was discussed in section 2.3, and the average NPL citation ratio does not correlate with the TIRs of the TDs (neither did any of the variants measuring NPL citations). Additionally, one of the most surprising results of the research was that the top 100 most cited patents had a lower NPL citation ratio than the total patent sets for every one of the 28 TDs. These observations do not mean that science does not contribute to the increase in technological performance, as there is far too much evidence to the contrary to seriously consider this option. One explanation is that the NPL is simply a poor measure of scientific reliance, and that unlike prior patents, the patent applicants and examiners are not highly motivated to ensure that all NPL is included in the citations. While there have been a number of studies that have used NPL ratio as a measure of science, perhaps the measure is not as reliable as previously thought.

A related explanation for the results is that the relationship between the basic science and the most important patents and the improvement in a domain are more complex than the NPL ratio tests. For example, the patents that rely very heavily on science may themselves not be incredibly impactful, but could lead to future patents (one or two citation generations later) that are very impactful. This case may be similar to the case of patents where the number of NPL citations is a poor measure of scientific impact, but rather a measure of scientific input and thus a measure that is more closely linked to the impact of a particular piece of non-patent literature may be more appropriate.

Another explanation for this could be that some aspect of the domain makes it easier or more difficult to assimilate science-based knowledge and thus the amount of NPL citations would not account for traits that are internal to the TD. Fleming and Sorenson (2004) explore the idea that technological development is influenced by science only when the technologies are very complex.

*'Science alters inventors' search processes, by leading them more directly to useful combinations, eliminating fruitless paths of research, and motivating them to continue even in the face of negative feedback' (Fleming and Sorenson, 2004)*

They results utilize patents and indicate that science is most useful when a technological domain involves many coupled components. They explain this relationship by saying that if the components of a technology are easily separated, then science offers less advantage, it is only when tightly coupled components cause significant challenges that the "map" provided by science is able to provide a significant benefit to the inventor.

Ultimately while this experiment does not show a significant relationship between NPL ratio and TIRs, it is unlikely that science does not play a role at all in technological development but rather that the NPL is an imperfect metric and/or that the connection between basic science and technological improvement is rather complex.

### 5.2.1.4. Recency: Citation age, patent age when cited and patent age

There are several strong links between the recency based DPMs and the TIRs, including DPMs that measure the publication date of the patents, the number of forward citations within 3 years of publication and the age of the backward citations. The three signals measure recency looking forward, backward and absolutely, and thus it is very likely that the recency of a technology is a strong explanatory factor in the variation of TIR's among the set of TD's. The idea that newer technologies improve more rapidly is consistent with many intuitive understandings about the progression of technology (the absolute measure); however, that the backward citation age and citations in the early years of a patent signal more rapid improvement has not been noted before.

A potential explanation for the correlation between the publication date the and TIRs is Darwinian in that the new TDs that do not improve more rapidly than the existing technologies never develop into anything important and thus are not seen in studies like this. If there are a large number of potential TDs being developed at all times, it is likely that only the TDs that improve more rapidly than the current state of the art will be developed further, and thus patented, diffused and studied by technological change researchers. This interpretation is also supported by the data that shows a wider spread of fast improving TIRs in more recent years than in the past, while there are some new domains that are improving very rapidly, there are also some slower improving domains that have recent average publication dates.

The correlation between the age of the backward citations and TIRs indicates that domains that improve more rapidly rely upon more recent patents. This could be due to the fact that fast improving domains make other knowledge obsolete more quickly, in that a technology is improving so rapidly that older components or knowledge are no longer useful to the domain of interest soon after they are created. Conversely, the knowledge that fast moving domains rely upon could be improving rapidly and thus the older knowledge and components become obsolete more quickly. The casual relationships are not clear from the citation-age correlation, but the fact that TIR-weighted backward citations are strongly correlated with TIRs may indicate that casual direction is from the fast moving components and not from the fast moving domain.

This is consistent with the forward citation based recency measure (CIT3) that states that in fast moving domains, patents are generally used significantly within 3 years after their publication. This measure is related to the patents within the domain (as opposed to the backward recency measure that is based mostly upon patents outside of the domain), it more clearly supports the casual direction of the base of knowledge from other domains becoming obsolete, in that when new knowledge is created it is used quickly and significantly often by other TD domains because the domain of interest is moving very quickly. In summary, if a domain is moving quickly, it is likely that the patents in that domain will be used within the first 3 years by other domains, after which the other domains will begin to rely more upon the next generation of patents. The casual relationship here is that the recency based measures are caused by the rapid improvement in the technological domain.

It is also possible that citations do not reflect usage but instead just competitive action, but this is not consistent with the fact that patent examiners force citations as a means of limiting claims and it is also not consistent with the evidence that citations do in fact relate to impact of a patent (Trajtenberg, 1990; Hall et al, 2001). We also note that the correlation of TIR with age of citations is independent of the date of emergence of the technology as shown by the forecast stability established in section 4.3.7. So even relatively "old" technologies like IC processors have maintained a high CIT3 throughout their life cycle. Nonetheless, there is a tendency for new domains to be advancing more rapidly. Therefore, the effect of recency is likely due to a combination of these theories but nonetheless is a powerful explanatory factor of TIRs.

### 5.2.1.5. Spillover: Knowledge Base Breadth and Vitality

The relationship between the breadth of knowledge based DPMs and TIRs was somewhat of a mixed signal, which is actually consistent with the prior literature in section 2.2. The results indicate that the use of knowledge that is external to the TD of interest is high in all domains and does not correlate highly with TIR. However, when weighted for the improvement rate of the cited TDs, the correlation with external knowledge use is quite high with the caveat that it may be less useful for future prediction rather than past understanding of technological improvement rates. This combination supports the idea that TDs that rely upon other TDs that are improving more rapidly are more likely to improve quickly themselves. This theory is intuitively consistent: if a domain is relying upon a slowly improving domain for certain components and those components are not improving rapidly, the improvement must come from more clever ways of integrating the same components and intuitively will be slower than another domain reliant upon rapidly improving components. For example, if an electric vehicle maker is relying upon electrochemical battery technology, then they are unlikely to improve very rapidly as the electrochemical battery TD has a relatively low TIR. On the contrary, if a domain is relying upon other technologies that are rapidly improving, such as integrated circuit processors, the improvement rate of the domain of interest is likely to improve even if very little is changed within the domain simply because the IC components that are used in the newer iterations are significantly better than past versions.

The correlation between citing fast improving domains and TIR is consistent with the recency casual links as mentioned in the previous section. A possible casual explanation is the domains that rely upon knowledge or components from fast moving domains will likely improve quickly themselves for the reasons mentioned above, and thus the knowledge derived from the fast moving domains will become obsolete more quickly as the cited domain is improving rapidly. A casual link in the other direction seems less likely, that fast moving domains will cause the other components that they are relying upon to improve more rapidly. While there is some theoretical support for this in that the enabling technologies will have to improve more rapidly to 'keep up' with the fast moving domain (Rycroft, 2006), this is not consistent with the wide variability of improvement rates of the domains that fast improving domains rely upon - that is: Why are some of the cited domain improving very rapidly and others improving very slowly even if they are both cited heavily by the fast improving domain?

The hypothesis about TIR weighted citation frequency is intuitive but it nonetheless is also apparently new. The theory that technological domains improve as a weighted sum of the improvements made in their knowledge base is logically appealing. The hypothesis that the rate of improvement of this knowledge base can be measured by a TIR-weighted sum of the domains they cite also has some logical basis. While the strong correlation of the TIR weighted backward citations with TIR supports this new theory, it must be noted that this measure is the most incomplete of all of the DPMs because a stringent test of this theory relies on knowing the TIRs of all of the TDs that are cited by a technology. As this study has only categorized 28 TDs that makeup ~10% of the patent database, it is possible that with the inclusion of all TDs that the theory would not be confirmed but it is also possible that it would be in this hypothetical case. If all of the patents were categorized and TIRs for each of the domains were determined (a probably impossible task since new domains constantly emerge), the TIR weighted backward citations could be even more reliable of a measure, but would still be incomplete due to the fact that both the patent sets and TIRs are subject to experimental error. Even with only 28 domains covered, it is likely that TDs that rely on knowledge from more rapidly improving TDs improve more quickly themselves, given the caveat that with a limited number of domains covered that a large portion of that value will come in the form of self-citations and relying on inventions within the domain.

#### 5.2.1.6. Domain Specific Technical Factors

While this research has approached the study of technological change from a socio-technical system point of view, there are a number of theories that point to domain specific technical factors as reasons for differing improvement rates.

One of the theories is that technological domains are highly sensitive to scaling effects and when a technology improves as it becomes smaller it is more likely to be rapidly improving than a TD that improves as it gets larger (i.e. integrated circuits vs combustion engines (Funk and Magee, 2014; Funk, 2011). Another theory is that improvement rates are largely determined by the number of interactions within a system, which is essentially a measure of complexity that is related to the number of components of a system that are affected when one is improved or changed (McNerney et al, 2011; Koh and Magee, 2008). If a large percentage of the system is affected whenever a single component is changed, it requires more work to integrate any new components in the system and thus the domain would improve more slowly. A third theory is that inventions can be classified into different hierarchy levels of change such as materials improvement (very low hierarchy level) or system operation change (very high hierarchy level) and that the distribution of these inventions is related to the improvement rate of a technology (Benson and Magee, 2012, 2014; Magee, 2012). Other ideas –that are not on the face consistent with the data –exist; for example that chemical based technologies are slower than purely electrical systems, in addition, it may be possible to develop other ideas based on technical differences among the domains for explaining the variation in rate of improvement.

It is quite likely that, while such domain specific technical factors were not explicitly discussed in this thesis, information about each specific TD might be extracted from the patents sets that have been collected for this research. This could be done using textual analysis and or more subjective interpretations of the particular domain. For example, if one was looking to evaluate the complexity of a TD, one could search the full text of the patent sets collected here for words such as 'tradeoff' or 'however', these types of experiments have been run at a first order level (Basnet, 2013) and are potential pathways forward for determining technology specific factors that may influence TIRs.

The improvement rates of technological domains are likely determined by a combination of system-level theories as were tested in this thesis and domain-specific factors that can be tested by further analysis of the patent sets.

#### 5.2.1.7. Summary of Direct Contributions of the Results of this Thesis

The results of this thesis provide a number of contributions to technological change theory, which are summarized in figure 92 below. The strongest link between TIRs and inventions are the frequency of recent and important inventions. There is a link between extra domain knowledge and TIRs that is applicable as long as it is weighted by the TIRs of the cited domains. There seems to be a negative relationship between non-patent-literature and the most important inventions. Finally the relationship between demand, revenue and patents has been shown to be nearly equal when measured against TIR using Sahal's relationship. The limitations of this technique as opposed to simply using the direct time variable have also been shown.

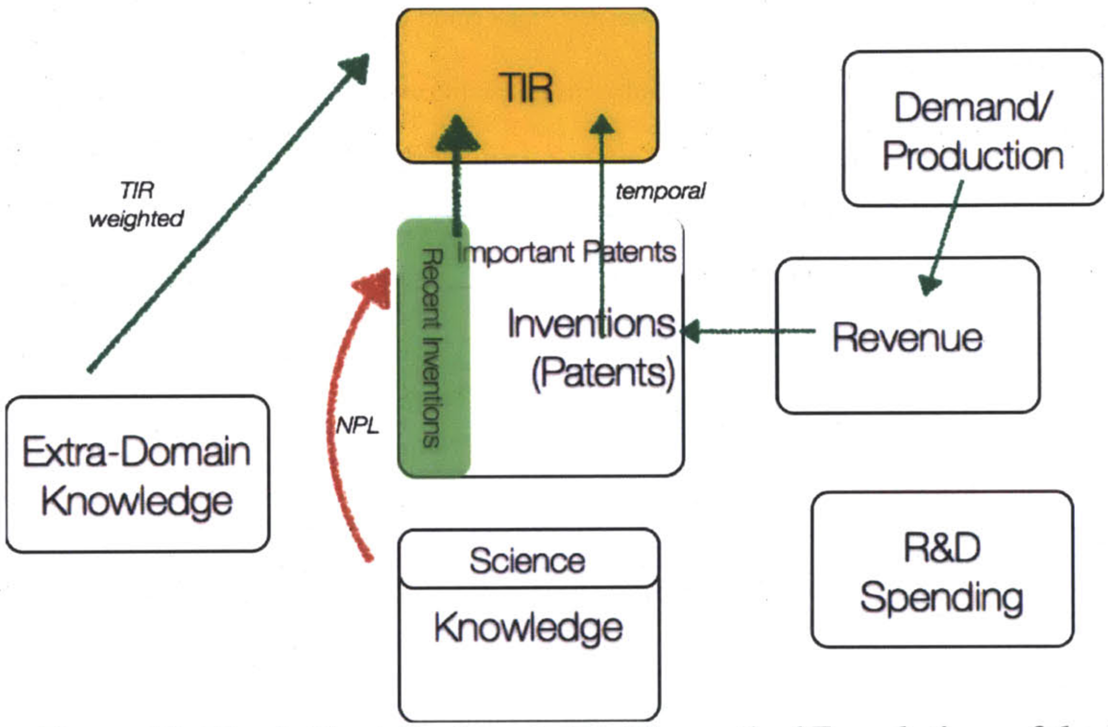



**Figure 92: Thesis Contributions to the Theoretical Foundations of the field of Technological Change**

### 5.2.2. Rising Sea Metaphor

As just discussed, the results of this research provide insight into a number of specific links in the complex web of technological change. There are many other links that were not explored in this research, but are consistent with the results that are shown here. An analogy that can be used to explain the complex interactions between the many different components of the highly complex socio-technical system of technological change is the 'Rising Sea Metaphor".

The Rising Sea metaphor is framed starting with a sea of knowledge which represents the body of knowledge from which all of technology is derived and represents pieces of knowledge such as patents, scientific journal articles and non-tracked knowledge (such as the knowledge that comes from a worker manufacturing a technology). Within this sea, each TD is represented by a region, these regions are not separated by clear distinct boundaries and overlap with many other regions in the same way that the patent sets in this study are overlapped with many other domains.

This sea is filled in two ways, one is by a well of science at the bottom of this sea which pumps water that represents scientific knowledge indiscriminately to all of the domains and represents an ever increasing source of knowledge from which future TDs can draw from and although it does not make any differentiated contributions to the improvement rates of specific TD's, it is ultimately the basis of progress in all domains. The other way that the sea of knowledge is filled is by the addition of patents or inventive knowledge that is selectively and intentionally poured onto the top of the sea by individuals, private companies and governments that are publishing patents regarding each particular TD.

Value can be extracted from this sea of knowledge by domain specific pumps far above the surface of the sea, and the TIR (improvement rate) of a TD is determined by the amount of "water" that can be extracted from the sea of knowledge. There are two main factors that impact the amount of water that can be extracted by a TD (and thus that impact the TIR): the level of the water in the sea that the pump is drawing from and the domain specific capability of the pump. In this analogy the level of the water represents the base of knowledge, which is being drawn from, and a higher level is better, while the capability of the pump represents the domain-specific technical factors that were not studied in great length in this thesis.

Attempts to adjust the water level that each TD draws by pouring more water into that region (represented by the addition of more patents and inventive effort in a particular TD) are not likely to be effective based on the metaphor since the water levels will quickly return to an equilibrium as the water level (base of knowledge) is ~90% determined by the patents (and thus inventive effort) in other technological domains plus generally available scientific knowledge pushing the % well beyond 90%. In this analogy, the addition of a large amount of additional domain-specific inventive effort does not significantly impact the TIR of a domain in the long term. While the addition of significantly more effort to one domain is not very impactful, the general distribution of inventive effort back into the sea of knowledge is approximately equal to the societal gains (revenue or profits) that can be extracted from the TIRs, this is to say that profits and demand are good indicators of where inventive effort is placed, but not necessarily the improvement rates of a technology.

Technological domains generally rely upon knowledge from outside their TD, and in this analogy the fast improving domains generally pull water from other knowledge bases that have high water levels and high TIRs.

Another way that this analogy fits with the results of this experiment are that the most recent inventions tend to be on the top of the sea because they were added most recently, and the fast improving domains are able to extract that recent knowledge more easily either due to where they are drawing their information (water) from or a better pump to extract the new information. One explanatory theory for why newer technologies tend to be more rapidly improving is that all technologies are linked together in a complex way and the exponential growth of each of these technologies creates a 'rising tide' that raises all boats at an exponential rate. This can be thought of like a single TIR for all of technology, and as newer TDs are discovered and utilized, they tend to be more rapidly improving because they are relying on general technology that is better than it was in the past due to the overall advancement of technology in general.

The sea in this analogy is continually rising at an accelerating rate due to the additions of science and patents into the sea proportional to the current sea level; this represents the general exponential improvement of technology as a whole. While the sea is rising, more and more pumps are continually being added to the pool in the hopes to find a new combination of knowledge from which to draw from and hopefully be able to extract large amounts of knowledge and economic/human value. The pumps that are less effective than the current pumps are unable to gather enough water to be as economically useful as a more effectively pumped domain - these represent the countless failed or surpassed technologies over the years.

Figure 93 shows a first order example of some of the characteristics of the Rising Sea analogy of technological change.

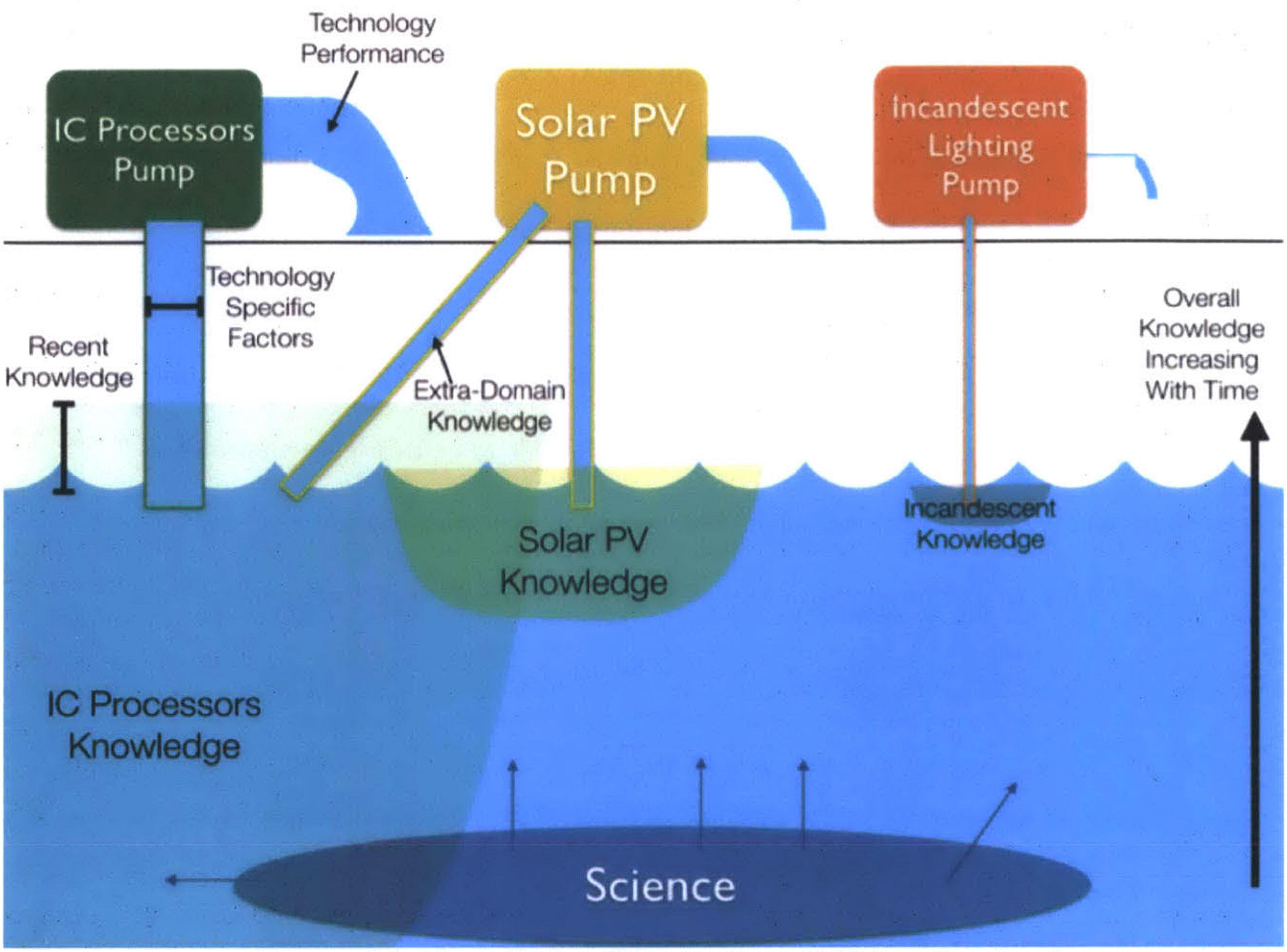



**Figure 93: First Order Illustration of Rising Sea Analogy**

## 5.3. Practical Implications

While the research adds to the academic theories of technological change, there are also immediate practical applications of the results of this study resulting from the models that can be used to predict the TIRs and the methodology for finding the correct set of patents to make the prediction. This section will discuss how the predictive models can be used to reduce technological uncertainty using only patent analysis. The section will also identify several groups of people who could benefit from use of the tools and understanding generated in this thesis.

### 5.3.1. Can Technological Improvement Rates be predicted?

The TIR of a domain can be very useful in understanding the potential of a certain technology particularly if one compares it to the TIR of competitive and complementary TDs. In short, TIRs essentially can help forecast the future of technology and entire industries. While this potential is very powerful, determining the TIR of even one domain can be very difficult, tedious and time consuming, and is often not possible depending on the availability of data. These issues are the main reason why TIRs have been found for only a small percentage of possible domains.

The results of the comparison of the DPM and TIRs revealed several markers with very strong correlations with the TIR. Additionally, these correlations were shown to be robust to the domains analyzed and consistent for 12 years into the future (2001-2013). As was mentioned previously the correlations do not provide good indications of casual direction, but nonetheless can be used for the forecasting of TIRs. It is likely that as the understanding of the relationship between the DPMs, the TIRs and the forecasted TIRs increases, the casual relationship will increase in turn. Nonetheless the DPMs are useful in forecasting even if they are not causes for the faster rates-they still statistically (in a robust way) reflect what is likely to happen –or at least what is happening now in performance trends. Thus the DPMs were combined into predictive models that allow for the estimation of TIRs given only a technological domain. The process of estimating a TIR given a domain works as follows:

1. Select a TD of interest
2. Use the COM to select a set of patents that represent the TD
3. Calculate the Cit3, AvepubYear of the patent set
4. Use the predictive model to estimate the TIR

TIR = -31.1968 + 0.1406 * Cit3 + 0.0155 * AvePubYear

The R2 of the most accurate predictive model is 0.64, which states that 64% of the variation in the TIR can be explained by the variation in the DPMs included. This actually is a positive sign for the model, as it is unlikely that all of the variation of the TIRs can be explained by the socio-technical system theories of technological change that were tested in this thesis. It is very likely that much of the rest of the variation in the TIRs is due to technology-specific factors (such as scaling or complexity) or to random variations (it is almost certainly impossible to explain 100% of the variation in TIRs).

These models could provide estimations for nearly any technological domain without the need of deep technical expertise and several man-months worth of effort. This predictive capability essentially opens up the capability for prediction of future technological capabilities to nearly any technological domain of interest.

## 5.3.2. Practical Benefit of Reduced Technological Uncertainty

There are many industries that would benefit from the reduction in technological uncertainty. Examples from several industries were given in section 1.1 of this thesis and will be discussed more here.

One of the main contributions of this thesis is the COM, and while it was designed to be used primarily for academic research, the method and its principles certainly have application elsewhere. The ability to quickly and easily select a set of patent related to a particular TD is useful for patent examiners, technology licensing offices at universities, patent attorneys, and individual inventors. The COM requires very little prior knowledge about the patent classification system to be used at a first-order level, and thus the inclusion of this method in patent searching tools could help make the patent system more accessible to a broad array of people. There are a number of other areas where specific documents or pieces of information are categorized using multiple classification systems such as legal documents, healthcare records, and government reports that could potentially be located using the same general classification overlap principles that are used in the COM.

The contributions of this thesis can be used by technological strategists for organizations that are interested in long-term planning of their technical capabilities. For example, a manufacturing company may be very interested in the relative improvement rates of traditional milling machines and that of SLA 3D printing. While 3D printing may not be a viable option for mass manufacturing of certain items at this point (although the author has used 3D printing as a production technique for a company that produces RFID-based rings), it is possible that it will become so in the future. The TIR of SLA 3D Printing can then provide a forecast of what the capabilities of 3D Printing technology will be in the future, potentially increasing to a point where a manufacturer would be interested in converting. Additionally, more in-depth analysis of the connectivity of component technologies to the improvement rate could also be used to guide the strategy of firms that want to advance component technologies, e.g., lasers or photopolymers for SLA 3D printing.

If the same manufacturer were more interested in metal-based 3D printing, the contributions of this thesis would still be helpful even though that particular domain was not studied. Using the predictive models and the COM, it would be possible for the TIRs to be estimated for a number of metal-based 3D printing technologies such as selective laser sintering or direct metal deposition. We believe that this ability to predict TIR using only patent information is an incredibly powerful tool that will allow for inexpensive estimation of TIR for a wide variety of domains that have not been studied before.

The reduction of technological uncertainty brought about by the TIRs and the predictive models are also useful for many other groups, most notably investors. Because manufacturing based TIRs may impact a great number of products that investors would consider, they are likely to be more valuable than the more product focused TIRs in the past. Additionally, the ability to quickly estimate a TIR for a new technology could provide a significant information advantage for an investor who is focused on technology-based companies. For example, an investor who was interested in a new type of 3D printing could first find the set of patents easily using the COM and then use the predictive models to estimate the TIR, which, if high, could spur an early investment in a rapidly improving industry.

The practical benefits of the newly created TIRs for manufacturing and energy technologies and the predict models for firms are relatively straightforward, there are important use cases for policy contributions as well. There are a number of government agencies (DOD, DOE) and private entities that attempt to spur the increase in capability of specific technologies. The approaches that are taken to do this vary widely but are generally based upon theories of technological change (Kazmerski, 2009). Such organizations should take note of the 'Rising Sea Metaphor' and understand the ways in which technologies interact with each other and how they change over time. In particular, we have shown that it is unlikely that the production function is the fundamental cause of technological change. In fact, we see that time is the more fundamental variable that technological progress should be measured against, which itself is not a very appealing answer. However, the hypotheses tested in this thesis and the theoretical contribution derived therefrom should provide indicators as to how specific technologies over time. Thus, while demand enhancement policies (subsidies, sales at non-market prices, etc...) are likely to prove ineffective at significantly increasing technological improvement rates for a particular domain, subsidizing R&D broadly does make great sense. Using solar PV as an example and returning to the 'Rising Sea Metaphor', the idea that a more effective 'pump' for PV progress would be achieved because of Silicon valley interest and significant increases in production of PV seems to have been an error and an avoidable one - to those who understand/believe that domain-specific technical factors and cumulative production are the root of causation for fast technical performance improvement.

# Chapter 6: Conclusions

This thesis was the first comparison between metrics that were derived from patent sets that represent a set technological domains (TD) and the performance improvement rates of those domains. This was done both to initiate predictive theory development and in order to quantitatively test hypotheses derived from existing qualitative theories of technological change.

A novel method for selecting sets of patents that represents a technological domain called the classification overlap method (COM) was used to select highly complete and relevant patents sets for each of the 28 TDs studied. A representative sampling of patents from each set were read to ensure the relevancy to the TDs, additionally a set of the top 100 most cited patents that were all related to the specific TD were collected.

Technological improvement rates were located for each of the 28 technological domains. Several of the TDs were measured using different FPMs resulting in a total of 72 domain-metric pairs for which TIRs were collected. Included in these 72 TIRs were the first examples of tracking the performance of manufacturing technological domains such as SLA 3D printing, milling machines and photolithography. A battery of statistical measures was developed to test the reliability of the TIRs, resulting in 42 reliable TIRs.

The most complete and reliable TIR for each TD was selected and compared with a set of domain-patent-markers that represented five hypotheses developed from theories of technologies change. There were weak or non-existent correlations with TIRs for the number of patents in a domain and the amount of non-patent literature cited by the patents in a domain. Stronger correlations with TIRs were shown for patent sets with a higher number of average forward citations and a more recent average publication date. The strongest correlations with TIR were the average number of forward citations within 3 years and the TIR weighted backward citation ratio, which measured the sources of information that a domain relies upon. The DPMs with the strongest correlation with TIR were tested for temporal and domain robustness and were then used to construct predictive models that are capable of explaining 64% of the variation in TIRs through measures derived from the patents.

The main theoretical findings of the thesis are in support of the theory that the number of patents is an indicator of inventive effort rather than inventive impact. Patents, revenue, and production can all be used as effort variables when compared with technological performance and have been shown to be equivalent for one domain according to Sahal's relationship, the results indicate however, that the most fundamental metric to compare with technological performance is time due to the large amount of influence a domain receives from other domains. The findings also suggest that the impact of very important singular inventions is overstated. The work does not support the simple theory that basic science is strongly linked to performance improvement in a specific technological domain. It was found that the use of more recent knowledge is in fact correlated with improved technology. Finally the research does not support the theory that more external spillover results in faster-improving technology, but rather that reliance specifically upon rapidly improving domains is important in achieving higher rates of technological improvement. These theoretical implications have been synthesized into one overarching analogy of technological change called the 'Rising Sea' metaphor.

## 6.1. Limitations and Future Work

The main limitations of this research are due to the use of patents as a proxy for inventions. These limitations were discussed at length in the prior art section of the thesis and still hold true for this research. Additionally, while we believe that the patent sets provided in this research represent some of the most complete and relevant sets ever constructed to represent such a wide variety of technological domains, there are undoubtedly missing patents from each TD and the relevancy of all of the sets are below 100%. The TIRs were also subjected to many different reliability tests, but it is still certain that each technological improvement curve is missing data points and there are omitted variables in each of the FPMs constructed to measure the performance of each TD. Finally, many of the results of this study are based upon correlations, which are not intended to represent casual links and thus care must be taken when interpreting the correlation values. The prediction models do provide good evidence of what change is currently happening and meaningful forecasts of the future within the specified robust time frame of 12 years, however past results are not always indicative of future returns and the estimations of the TIRs are subject to the same disclaimer.

The main goal of this research is to encourage the trend of creating falsifiable tests for theories of technological change in order to create a truly cumulative field of work where the average technological change paper is cited very heavily within the first 3 years of its publication. With this in mind, future work that builds upon this thesis should at all times continue to keep in mind the scientific principles of repeatability and the ability to falsify. The addition of more domains should be included in any future study, while 28 domains represents a significant start, there are still many more technical areas that should be added and analyzed using the same tools shown in this thesis. Additionally, while over 100 FPMs were used in evaluating the five hypotheses, there are certainly more that can be developed and tested - although it must be noted that additional complexity often times comes at the result of ease of understanding, which should be at the core of any heuristic used to test a hypothesis. Furthermore, as 'big data' tools continue to be developed and refined it may one day be possible to perform far more rigorous tests of the patents by using natural language processing, machine learning and crowdsourcing to locate and analyze large sets of patents to better understand how and why technology changes over time. Finally, as more time passes and more TDs are analyzed using tests such as the ones presented in this thesis, the results of the thesis should be checked for accuracy in the future.

# Appendices

## Appendix A: TIR Reliability Measures

This appendix displays the statistical reliability measures for each of the 72 DMPs split up into the 4 different subsets. Each table displays the DMP in the first column, followed by the TIR (k-value) the number of data points (n), the R2 value, the confidence interval Standard Deviation (stdev) and the Point Removal Method standard deviation (stdevPRM) in the 6th and final column. Table table A.1 shows the values for all data points and all years.

| DMP | AllAll k | AllAll N | AllAll R2 | AllAll stdev | AllAll stdevPRM |
|---|---|---|---|---|---|
| 3D-Printing (industrial stereolithography) (mm3/sec*$) | 33% | 5 | 0.7 | 14% | 15% |
| 3D-Printing (industrial stereolithography) (1/sec*$ (including build volume/machine size)) | 37.60% | 5 | 0.92 | 6.70% | 6.30% |
| 3D-Printing (industrial stereolithography) (1/sec*$) | 30.60% | 7 | 0.9 | 4.20% | 1.30% |
| 3D-Printing (industrial stereolithography) (1/sec) | 34.20% | 6 | 0.89 | 5.70% | 3.60% |
| Aircraft Transport (passenger*mph) | 12.20% | 12 | 0.98 | 0.60% | 0.20% |
| Camera Sensitivity (mV/micron2) | 15.60% | 11 | 0.99 | 0.50% | 0.20% |
| Capacitor Energy Storage (W*hr/$) | 18.30% | 7 | 0.82 | 4.10% | 1.80% |
| Capacitor Energy Storage (W*hr/kg) | 14.10% | 12 | 0.96 | 0.90% | 0.40% |
| Capacitor Energy Storage (W*hr/L) | 14.40% | 14 | 0.73 | 2.40% | 1.10% |
| Combustion Engines (W/$) | 2.40% | 79 | 0.18 | 0.60% | 0.10% |
| Combustion Engines (W/kg) | 3.60% | 99 | 0.49 | 0.40% | 0.00% |
| Combustion Engines (W/L) | 0.80% | 127 | 0.05 | 0.30% | 0.00% |
| Combustion Engines - Aircraft (W/$) | 0.60% | 23 | 0.07 | 0.50% | 0.10% |
| Combustion Engines - Aircraft (W/kg) | 3.10% | 42 | 0.85 | 0.20% | 0.00% |
| Combustion Engines - Aircraft (W/L) | 2.10% | 67 | 0.32 | 0.40% | 0.00% |
| Combustion Engines - Aircraft Piston (W/$) | 2.70% | 13 | 0.2 | 1.70% | 1.30% |
| Combustion Engines - Aircraft | 6.80% | 21 | 0.8 | 0.80% | 0.20% |

| | | | | | |
|---|---|---|---|---|---|
| Piston (W/kg) | | | | | |
| Combustion Engines - Aircraft Piston (W/L) | 5.20% | 45 | 0.43 | 0.90% | 0.10% |
| Combustion Engines - Aircraft Turbine (W/$) | 2.30% | 10 | 0.27 | 1.50% | 0.60% |
| Combustion Engines - Aircraft Turbine (W/kg) | 0.70% | 21 | 0.18 | 0.40% | 0.10% |
| Combustion Engines - Aircraft Turbine (W/L) | 4.00% | 22 | 0.59 | 0.80% | 0.20% |
| Combustion Engines - Automobile (W/$) | 3.60% | 56 | 0.57 | 0.40% | 0.40% |
| Combustion Engines - Automobile (W/kg) | 3.20% | 57 | 0.74 | 0.30% | 0.10% |
| Combustion Engines - Automobile (W/L) | 1.10% | 61 | 0.47 | 0.10% | 0.00% |
| Computed Tomography (CT) (1/mm*sec) | 36.70% | 13 | 0.78 | 4.50% | 1.80% |
| Electric Motors (W/kg) | 2.90% | 13 | 0.83 | 0.40% | 0.20% |
| Electric Motors (W/L) | 2.60% | 16 | 0.95 | 0.20% | 0.10% |
| Electrical Energy Transmission (W*km/$) | 2.10% | 29 | 0.36 | 0.50% | 0.10% |
| Electrical Energy Transmission (W*km) | 12.60% | 33 | 0.88 | 0.80% | 0.20% |
| Electrical Energy Transmission - AC (W*km/$) | 1.30% | 9 | 0.14 | 1.50% | 0.50% |
| Electrical Energy Transmission - AC (W*km) | 12.60% | 10 | 0.86 | 1.80% | 0.60% |
| Electrical Energy Transmission - DC (W*km/$) | -2.10% | 20 | 0.08 | 1.80% | 0.30% |
| Electrical Energy Transmission - DC (W*km) | 0.30% | 23 | 0 | 2.90% | 0.70% |
| Electrical Information Transmission (kB/sec*$) | 14.30% | 10 | 0.9 | 1.70% | 0.90% |
| Electrical Information Transmission (kB/sec) | 13.50% | 12 | 0.9 | 1.30% | 0.50% |
| Electrochemical Battery Energy Storage (W*hr/$) | -3.30% | 31 | 0.13 | 1.70% | 0.40% |
| Electrochemical Battery Energy Storage (W*hr/kg) | 2.40% | 64 | 0.62 | 0.20% | 0.00% |
| Electrochemical Battery Energy Storage (W*hr/L) | 3.30% | 47 | 0.77 | 0.30% | 0.00% |
| Electronic Computation (1/sec) | 33.90% | 29 | 0.91 | 1.50% | 0.40% |
| Electronic Computation (1/sec*$) | 40.40% | 80 | 0.93 | 0.90% | 0.20% |
| Electronic Computation (1/sec) | 34.40% | 92 | 0.67 | 1.80% | 0.30% |
| Flywheel Energy Storage (kWhr/kg) | 9.30% | 8 | 0.9 | 1.40% | 0.90% |
| Fuel Cell Energy Production (kW/$) | 14.70% | 9 | 1 | 0.30% | 0.20% |
| Genome Sequencing (1/$) | 29.80% | 8 | 0.96 | 2.10% | 1.00% |
| Incandescent Artificial Illumination (1000*lm*hr/$) | 2.70% | 11 | 0.6 | 0.80% | 0.40% |
| Integrated Circuit Information Storage (transistors/die) | 43.20% | 20 | 0.98 | 0.90% | 0.30% |
| Integrated Circuit Processors (transistors/die) | 36.30% | 12 | 0.97 | 1.60% | 0.70% |
| LED Artificial Illumination (lm/$) | 24.40% | 13 | 0.98 | 0.80% | 0.30% |
| LED Artificial Illumination (lm) | 35.80% | 17 | 0.98 | 0.90% | 0.40% |
| Magnet Resonance Imaging (MRI) (1/mm*sec*$) | 47.50% | 6 | 0.88 | 7.40% | 11.30% |
| Magnet Resonance Imaging (MRI) (1/mm*sec) | 21.30% | 11 | 0.86 | 2.70% | 0.90% |
| Magnetic Information Storage (megabits/$) | 27.50% | 33 | 0.87 | 1.50% | 0.40% |
| Magnetic Information Storage | 27.00% | 38 | 0.89 | 1.20% | 0.40% |

| | | | | | |
|---|---|---|---|---|---|
| (megabits/cc) | | | | | |
| Magnetic Information Storage - Hard Disk (megabits/$) | 31.80% | 19 | 0.89 | 2.00% | 0.70% |
| Magnetic Information Storage - Hard Disk (megabits/cc) | 32.70% | 24 | 0.96 | 1.00% | 0.30% |
| Magnetic Information Storage - Tape (megabits/$) | 24.60% | 14 | 0.84 | 2.70% | 1.00% |
| Magnetic Information Storage - Tape (megabits/cc) | 23.80% | 14 | 0.92 | 1.70% | 0.60% |
| Milling Machines (hp/mm) | 3.50% | 8 | 0.79 | 0.90% | 3.90% |
| Optical Information Storage (Mb/$) | 34.40% | 4 | 0.7 | 24.00% | 19.20% |
| Optical Information Storage (Mb/cc) | 27.20% | 16 | 0.95 | 1.40% | 0.60% |
| Optical Information Transmission (Kb/sec*km*$) | 63.00% | 17 | 0.92 | 2.70% | 1.70% |
| Optical Information Transmission (Kb/sec) | 66.40% | 17 | 0.88 | 3.60% | 2.20% |
| Photolithography (1/micron*$) | -5.60% | 10 | 0.59 | 2.00% | 0.90% |
| Photolithography (in2/hr*micron*$) | 5.30% | 6 | 0.21 | 6.50% | 3.30% |
| Photolithography (in2/hr*micron) | 24.00% | 11 | 0.85 | 3.00% | 0.80% |
| Photolithography (in2/hr*$) | -3.60% | 6 | 0.19 | -1.457 | 1.20% |
| Solar Photovoltaic Energy Generation (kWhr/$) | 10.80% | 34 | 0.98 | 0.30% | 0.10% |
| Solar Photovoltaic Energy Generation (W/$) | 9.0% | 9.8% | 8.6% | 9% | 35 |
| Superconductivity (1/K) | 3.80% | 15 | 0.87 | 0.40% | 0.10% |
| Wind Turbine Energy Generation (W/$) | 2.90% | 23 | 0.67 | 0.50% | 0.10% |
| Wireless Information Transmission (bits/sec*m2) | 31.30% | 19 | 0.77 | 3.20% | 0.60% |
| Wireless Information Transmission (bits/sec*Hz) | 14.40% | 42 | 0.92 | 0.60% | 0.10% |
| Wireless Information Transmission (kb/sec) | 17.10% | 49 | 0.64 | 1.60% | 0.60% |

**Table A.1: Statistical Measures for the Technological Improvement Rates of the 72 Domain-Metrics Pairs**

Table A.2 shows the statistical measures for just the non-dominated data points.

| **DMP** | AllND k | AllND N | AllND R2 | AllND stdev | AllND stdevPRM |
|---|---|---|---|---|---|
| 3D-Printing (industrial stereolithography) (mm3/sec*$) | 37% | 3 | 95% | 39% | 15% |
| 3D-Printing (industrial stereolithography) (1/sec*$ (including build volume/machine size)) | 37.30% | 4 | 0.93 | 9.70% | 6.40% |
| 3D-Printing (industrial stereolithography) (1/sec*$) | 27.30% | 4 | 0.97 | 4.90% | 4.50% |
| 3D-Printing (industrial stereolithography) (1/sec) | 28.50% | 4 | 0.92 | 8.80% | 4.00% |
| Aircraft Transport (passenger*mph) | 12.20% | 12 | 0.98 | 0.60% | 0.20% |
| Camera Sensitivity (mV/micron2) | 15.60% | 11 | 0.99 | 0.50% | 0.20% |
| Capacitor Energy Storage (W*hr/$) | 21.10% | 6 | 0.97 | 2.00% | 1.30% |

| | | | | | |
|---|---|---|---|---|---|
| Capacitor Energy Storage (W*hr/kg) | 15.00% | 11 | 0.98 | 0.60% | 0.30% |
| Capacitor Energy Storage (W*hr/L) | 16.50% | 7 | 0.72 | 4.90% | 2.70% |
| Combustion Engines (W/$) | 7.90% | 18 | 0.79 | 1.00% | 0.50% |
| Combustion Engines (W/kg) | 5.70% | 24 | 0.82 | 0.60% | 0.20% |
| Combustion Engines (W/L) | 3.50% | 14 | 0.56 | 1.00% | 0.90% |
| Combustion Engines - Aircraft (W/$) | 1.90% | 6 | 0.64 | 0.90% | 0.30% |
| Combustion Engines - Aircraft (W/kg) | 3.40% | 18 | 0.9 | 0.30% | 0.10% |
| Combustion Engines - Aircraft (W/L) | 3.50% | 14 | 0.56 | 1.00% | 0.90% |
| Combustion Engines - Aircraft Piston (W/$) | 10.20% | 4 | 0.93 | 3.50% | 3.00% |
| Combustion Engines - Aircraft Piston (W/kg) | 6.40% | 8 | 0.8 | 1.50% | 1.00% |
| Combustion Engines - Aircraft Piston (W/L) | 6.50% | 13 | 0.79 | 1.00% | 0.40% |
| Combustion Engines - Aircraft Turbine (W/$) | 2.90% | 7 | 0.79 | 0.80% | 0.30% |
| Combustion Engines - Aircraft Turbine (W/kg) | 1.90% | 10 | 0.98 | 0.10% | 0.00% |
| Combustion Engines - Aircraft Turbine (W/L) | 5.40% | 12 | 0.69 | 1.20% | 0.80% |
| Combustion Engines - Automobile (W/$) | 7.90% | 18 | 0.79 | 1.00% | 1.00% |
| Combustion Engines - Automobile (W/kg) | 5.60% | 17 | 0.85 | 0.60% | 0.30% |
| Combustion Engines - Automobile (W/L) | 1.40% | 27 | 0.88 | 0.10% | 0.00% |
| Computed Tomography (CT) (1/mm*sec) | 36.70% | 13 | 0.78 | 4.50% | 1.80% |
| Electric Motors (W/kg) | 3.10% | 11 | 0.84 | 0.50% | 0.20% |
| Electric Motors (W/L) | 2.60% | 15 | 0.94 | 0.20% | 0.10% |
| Electrical Energy Transmission (W*km/$) | 3.60% | 7 | 0.71 | 1.20% | 0.60% |
| Electrical Energy Transmission (W*km) | 14.90% | 10 | 0.97 | 0.90% | 0.40% |
| Electrical Energy Transmission - AC (W*km/$) | 5.80% | 4 | 0.53 | 7.10% | 10.70% |
| Electrical Energy Transmission - AC (W*km) | 14.90% | 6 | 0.92 | 2.50% | 2.00% |
| Electrical Energy Transmission - DC (W*km/$) | 6.20% | 5 | 0.92 | 1.50% | 0.70% |
| Electrical Energy Transmission - DC (W*km) | 7.90% | 5 | 0.74 | 3.70% | 3.90% |
| Electrical Information Transmission (kB/sec*$) | 14.30% | 10 | 0.9 | 1.70% | 0.90% |
| Electrical Information Transmission (kB/sec) | 13.70% | 11 | 0.9 | 1.40% | 0.60% |
| Electrochemical Battery Energy Storage (W*hr/$) | 3.10% | 11 | 0.77 | 0.60% | 0.50% |
| Electrochemical Battery Energy Storage (W*hr/kg) | 2.80% | 21 | 0.92 | 0.20% | 0.00% |
| Electrochemical Battery Energy Storage (W*hr/L) | 3.60% | 21 | 0.91 | 0.30% | 0.10% |
| Electronic Computation (1/sec) | 33.00% | 19 | 0.9 | 2.00% | 0.70% |
| Electronic Computation (1/sec*$) | 37.80% | 23 | 0.96 | 1.30% | 0.40% |
| Electronic Computation (1/sec) | 41.40% | 22 | 0.98 | 0.90% | 0.30% |
| Flywheel Energy Storage (kWhr/kg) | 9.00% | 7 | 0.92 | 1.40% | 1.10% |

| | | | | | |
|---|---|---|---|---|---|
| Fuel Cell Energy Production (kW/$) | 14.70% | 9 | 1 | 0.30% | 0.20% |
| Genome Sequencing (1/$) | 29.80% | 8 | 0.96 | 2.10% | 1.00% |
| Incandescent Artificial Illumination (1000*lm*hr/$) | 4.50% | 9 | 0.93 | 0.50% | 0.30% |
| Integrated Circuit Information Storage (transistors/die) | 43.20% | 20 | 0.98 | 0.90% | 0.30% |
| Integrated Circuit Processors (transistors/die) | 36.30% | 12 | 0.97 | 1.60% | 0.70% |
| LED Artificial Illumination (lm/$) | 24.40% | 13 | 0.98 | 0.80% | 0.30% |
| LED Artificial Illumination (lm) | 35.80% | 17 | 0.98 | 0.90% | 0.40% |
| Magnet Resonance Imaging (MRI) (1/mm*sec*$) | 47.50% | 6 | 0.88 | 7.40% | 11.30% |
| Magnet Resonance Imaging (MRI) (1/mm*sec) | 21.00% | 10 | 0.84 | 3.00% | 1.10% |
| Magnetic Information Storage (megabits/$) | 26.60% | 20 | 0.85 | 2.10% | 0.70% |
| Magnetic Information Storage (megabits/cc) | 24.00% | 28 | 0.89 | 1.30% | 0.30% |
| Magnetic Information Storage - Hard Disk (megabits/$) | 31.90% | 17 | 0.89 | 2.20% | 0.80% |
| Magnetic Information Storage - Hard Disk (megabits/cc) | 32.70% | 24 | 0.96 | 1.00% | 0.30% |
| Magnetic Information Storage - Tape (megabits/$) | 24.60% | 12 | 0.83 | 3.10% | 1.10% |
| Magnetic Information Storage - Tape (megabits/cc) | 23.80% | 14 | 0.92 | 1.70% | 0.60% |
| Milling Machines (hp/mm) | 3.40% | 6 | 0.96 | 0.40% | 3.60% |
| Optical Information Storage (Mb/$) | 34.40% | 4 | 0.7 | 24.00% | 19.20% |
| Optical Information Storage (Mb/cc) | 27.10% | 15 | 0.95 | 1.50% | 0.70% |
| Optical Information Transmission (Kb/sec*km*$) | 65.10% | 13 | 0.93 | 3.00% | 2.10% |
| Optical Information Transmission (Kb/sec) | 68.40% | 12 | 0.9 | 4.00% | 2.90% |
| Photolithography (1/micron*$) | 0.00% | 0 | 1 | #NUM! | #VALUE! |
| Photolithography (in2/hr*micron*$) | 9.10% | 2 | 1 | #NUM! | 0.00% |
| Photolithography (in2/hr*micron) | 25.20% | 7 | 0.97 | 2.00% | 1.30% |
| Photolithography (in2/hr*$) | 2.90% | 2 | 1 | #NUM! | 0.00% |
| Solar Photovoltaic Energy Generation (kWhr/$) | 10.80% | 34 | 0.98 | 0.30% | 0.10% |
| Solar Photovoltaic Energy Generation (W/$) | 9.20% | 34 | 29 | 0.94 | 0.95 |
| Superconductivity (1/K) | 3.70% | 14 | 0.86 | 0.50% | 0.10% |
| Wind Turbine Energy Generation (W/$) | 3.70% | 9 | 0.68 | 1.10% | 1.90% |
| Wireless Information Transmission (bits/sec*m2) | 35.00% | 10 | 0.92 | 2.90% | 1.40% |
| Wireless Information Transmission (bits/sec*Hz) | 15.40% | 19 | 0.96 | 0.70% | 0.30% |
| Wireless Information Transmission (kb/sec) | 18.10% | 18 | 0.67 | 2.80% | 2.90% |

**Table A.2: Statistical Measures for the Non-Dominated Technological Improvement Rates of the 72 Domain-Metrics Pairs**

Table A.3 shows the statistical measures for all of the data points since 1970.

| **DMP** | 70All k | 70All N | 70All R2 | 70All stdev | 70All stdevPRM |
|---|---|---|---|---|---|
| 3D-Printing (industrial stereolithography) (mm3/sec*$) | 33% | 5 | 0.7 | 14% | 15% |
| 3D-Printing (industrial stereolithography) (1/sec*$ (including build volume/machine size)) | 37.60% | 5 | 0.92 | 6.70% | 6.30% |
| 3D-Printing (industrial stereolithography) (1/sec*$) | 30.60% | 7 | 0.9 | 4.20% | 1.30% |
| 3D-Printing (industrial stereolithography) (1/sec) | 34.20% | 6 | 0.89 | 5.70% | 3.60% |
| Aircraft Transport (passenger*mph) | 8.40% | 2 | 1 | #NUM! | 0.00% |
| Camera Sensitivity (mV/micron2) | 15.60% | 11 | 0.99 | 0.50% | 0.20% |
| Capacitor Energy Storage (W*hr/$) | 18.30% | 7 | 0.82 | 4.10% | 1.80% |
| Capacitor Energy Storage (W*hr/kg) | 13.40% | 10 | 0.92 | 1.40% | 0.50% |
| Capacitor Energy Storage (W*hr/L) | 17.90% | 9 | 0.58 | 5.60% | 4.40% |
| Combustion Engines (W/$) | -4.40% | 25 | 0.11 | 2.90% | 0.70% |
| Combustion Engines (W/kg) | 3.50% | 37 | 0.07 | 2.10% | 0.30% |
| Combustion Engines (W/L) | 1.70% | 40 | 0.03 | 1.50% | 0.20% |
| Combustion Engines - Aircraft (W/$) | 2.50% | 7 | 0.15 | 3.30% | 1.40% |
| Combustion Engines - Aircraft (W/kg) | 0.30% | 18 | 0.02 | 0.50% | 0.20% |
| Combustion Engines - Aircraft (W/L) | 1.90% | 16 | 0.28 | 0.90% | 0.20% |
| Combustion Engines - Aircraft Piston (W/$) | #VALUE! | 0 | #VALUE! | 0.00% | 0.00% |
| Combustion Engines - Aircraft Piston (W/kg) | #VALUE! | 0 | #VALUE! | 0.00% | 0.00% |
| Combustion Engines - Aircraft Piston (W/L) | #VALUE! | 0 | #VALUE! | 0.00% | 0.00% |
| Combustion Engines - Aircraft Turbine (W/$) | 2.50% | 7 | 0.15 | 3.30% | 1.40% |
| Combustion Engines - Aircraft Turbine (W/kg) | 0.30% | 18 | 0.02 | 0.50% | 0.20% |
| Combustion Engines - Aircraft Turbine (W/L) | 1.90% | 16 | 0.28 | 0.90% | 0.20% |
| Combustion Engines - Automobile (W/$) | 0.70% | 18 | 0.03 | 1.00% | 1.00% |
| Combustion Engines - Automobile (W/kg) | 1.50% | 19 | 0.21 | 0.70% | 0.20% |
| Combustion Engines - Automobile (W/L) | 2.70% | 24 | 0.57 | 0.50% | 0.10% |
| Computed Tomography (CT) (1/mm*sec) | 36.70% | 13 | 0.78 | 4.50% | 1.80% |
| Electric Motors (W/kg) | 6.10% | 4 | 0.88 | 2.90% | 1.40% |
| Electric Motors (W/L) | 0.00% | 5 | 1 | 0.00% | 0.00% |
| Electrical Energy Transmission (W*km/$) | -3.40% | 19 | 0.08 | 3.20% | 0.60% |
| Electrical Energy Transmission (W*km) | -4.10% | 22 | 0.05 | 4.60% | 0.80% |
| Electrical Energy Transmission - AC (W*km/$) | 0.00% | 0 | 1 | #NUM! | #VALUE! |

| | | | | | |
|---|---|---|---|---|---|
| Electrical Energy Transmission - AC (W*km) | 0.00% | 0 | 1 | #NUM! | #VALUE! |
| Electrical Energy Transmission - DC (W*km/$) | -3.90% | 18 | 0.13 | 2.70% | 0.50% |
| Electrical Energy Transmission - DC (W*km) | -5.00% | 21 | 0.09 | 4.10% | 0.80% |
| Electrical Information Transmission (kB/sec*$) | 52.90% | 2 | 1 | #NUM! | 0.00% |
| Electrical Information Transmission (kB/sec) | 46.00% | 2 | 1 | #NUM! | 0.00% |
| Electrochemical Battery Energy Storage (W*hr/$) | -5.90% | 29 | 0.2 | 2.50% | 0.40% |
| Electrochemical Battery Energy Storage (W*hr/kg) | 6.00% | 53 | 0.5 | 0.80% | 0.10% |
| Electrochemical Battery Energy Storage (W*hr/L) | 7.10% | 37 | 0.53 | 1.10% | 0.20% |
| Electronic Computation (1/sec) | 40.50% | 17 | 0.76 | 4.20% | 1.70% |
| Electronic Computation (1/sec*$) | 51.30% | 53 | 0.92 | 1.30% | 0.40% |
| Electronic Computation (1/sec) | 53.10% | 65 | 0.65 | 3.00% | 0.70% |
| Flywheel Energy Storage (kWhr/kg) | 9.30% | 8 | 0.9 | 1.40% | 0.90% |
| Fuel Cell Energy Production (kW/$) | 14.40% | 5 | 0.99 | 0.90% | 0.60% |
| Genome Sequencing (1/$) | 29.30% | 7 | 0.91 | 3.80% | 3.80% |
| Incandescent Artificial Illumination (1000*lm*hr/$) | -2.90% | 2 | 1 | #NUM! | 0.00% |
| Integrated Circuit Information Storage (transistors/die) | 37.20% | 14 | 1 | 0.60% | 0.20% |
| Integrated Circuit Processors (transistors/die) | 36.30% | 12 | 0.97 | 1.60% | 0.70% |
| LED Artificial Illumination (lm/$) | 24.40% | 13 | 0.98 | 0.80% | 0.30% |
| LED Artificial Illumination (lm) | 36.20% | 15 | 0.97 | 1.20% | 0.60% |
| Magnet Resonance Imaging (MRI) (1/mm*sec*$) | 47.50% | 6 | 0.88 | 7.40% | 11.30% |
| Magnet Resonance Imaging (MRI) (1/mm*sec) | 45.70% | 7 | 0.85 | 6.80% | 12.10% |
| Magnetic Information Storage (megabits/$) | 38.20% | 26 | 0.85 | 2.30% | 0.90% |
| Magnetic Information Storage (megabits/cc) | 32.30% | 28 | 0.9 | 1.60% | 0.50% |
| Magnetic Information Storage - Hard Disk (megabits/$) | 39.20% | 17 | 0.86 | 3.00% | 3.50% |
| Magnetic Information Storage - Hard Disk (megabits/cc) | 36.10% | 19 | 0.94 | 1.70% | 0.70% |
| Magnetic Information Storage - Tape (megabits/$) | 36.60% | 9 | 0.85 | 4.80% | 2.20% |
| Magnetic Information Storage - Tape (megabits/cc) | 29.60% | 9 | 0.9 | 3.20% | 0.90% |
| Milling Machines (hp/mm) | 4.40% | 4 | 0.87 | 2.20% | 5.80% |
| Optical Information Storage (Mb/$) | 34.40% | 4 | 0.7 | 24.00% | 19.20% |
| Optical Information Storage (Mb/cc) | 27.20% | 16 | 0.95 | 1.40% | 0.60% |
| Optical Information Transmission (Kb/sec*km*$) | 63.00% | 17 | 0.92 | 2.70% | 1.70% |
| Optical Information Transmission (Kb/sec) | 66.40% | 17 | 0.88 | 3.60% | 2.20% |
| Photolithography (1/micron*$) | -3.50% | 9 | 0.43 | 1.90% | 0.70% |
| Photolithography (in2/hr*micron*$) | 11.60% | 5 | 0.44 | 10.30% | 6.70% |
| Photolithography (in2/hr*micron) | 28.50% | 9 | 0.78 | 5.10% | 2.10% |
| Photolithography (in2/hr*$) | -3.20% | 5 | 0.07 | -3.24 | 5.70% |

| | | | | | |
|---|---|---|---|---|---|
| Solar Photovoltaic Energy Generation (kWhr/$) | 10.80% | 34 | 0.98 | 0.30% | 0.10% |
| Solar Photovoltaic Energy Generation (W/$) | 0.93 | 0.94 | 0.4% | 0.4% | 0.4% |
| Superconductivity (1/K) | 9.10% | 8 | 0.74 | 2.40% | 0.70% |
| Wind Turbine Energy Generation (W/$) | 3.40% | 21 | 0.38 | 1.00% | 0.30% |
| Wireless Information Transmission (bits/sec*m2) | 47.30% | 16 | 0.42 | 10.10% | 3.00% |
| Wireless Information Transmission (bits/sec*Hz) | 14.10% | 39 | 0.67 | 1.40% | 0.30% |
| Wireless Information Transmission (kb/sec) | 40.00% | 45 | 0.69 | 2.80% | 0.50% |

**Table A.3: Statistical Measures for the Technological Improvement Rates of the 72 Domain-Metrics Pairs Since 1970**

Finally, Table A.4 shows the statistical measures for the non-dominated data points since 1970.

| **DMP** | 70ND k | 70ND N | 70ND R2 | 70ND stdev | 70ND stdevPRM |
|---|---|---|---|---|---|
| 3D-Printing (industrial stereolithography) (mm3/sec*$) | 37% | 3 | 0.95 | 39% | 15% |
| 3D-Printing (industrial stereolithography) (1/sec*$ (including build volume/machine size)) | 37.30% | 4 | 0.93 | 9.70% | 6.40% |
| 3D-Printing (industrial stereolithography) (1/sec*$) | 27.30% | 4 | 0.97 | 4.90% | 4.50% |
| 3D-Printing (industrial stereolithography) (1/sec) | 28.50% | 4 | 0.92 | 8.80% | 4.00% |
| Aircraft Transport (passenger*mph) | 8.40% | 2 | 1 | #NUM! | 0.00% |
| Camera Sensitivity (mV/micron2) | 15.60% | 11 | 0.99 | 0.50% | 0.20% |
| Capacitor Energy Storage (W*hr/$) | 21.10% | 6 | 0.97 | 2.00% | 1.30% |
| Capacitor Energy Storage (W*hr/kg) | 14.60% | 9 | 0.97 | 1.00% | 0.40% |
| Capacitor Energy Storage (W*hr/L) | 23.20% | 5 | 0.68 | 11.00% | 10.00% |
| Combustion Engines (W/$) | 0.00% | 0 | 1 | #NUM! | #VALUE! |
| Combustion Engines (W/kg) | 3.70% | 10 | 0.37 | 1.90% | 0.80% |
| Combustion Engines (W/L) | 4.10% | 8 | 0.45 | 2.10% | 1.70% |
| Combustion Engines - Aircraft (W/$) | 5.40% | 6 | 0.93 | 1.00% | 0.30% |
| Combustion Engines - Aircraft (W/kg) | 1.70% | 8 | 0.98 | 0.10% | 0.00% |
| Combustion Engines - Aircraft (W/L) | 2.30% | 7 | 0.8 | 0.60% | 0.60% |
| Combustion Engines - Aircraft Piston (W/$) | #VALUE! | 0 | #VALUE! | 0.00% | 0.00% |
| Combustion Engines - Aircraft Piston (W/kg) | #VALUE! | 0 | #VALUE! | 0.00% | 0.00% |

| | | | | | |
|---|---|---|---|---|---|
| Combustion Engines - Aircraft Piston (W/L) | #VALUE! | 0 | #VALUE! | 0.00% | 0.00% |
| Combustion Engines - Aircraft Turbine (W/$) | 5.40% | 6 | 0.93 | 1.00% | 0.30% |
| Combustion Engines - Aircraft Turbine (W/kg) | 1.70% | 8 | 0.98 | 0.10% | 0.00% |
| Combustion Engines - Aircraft Turbine (W/L) | 2.30% | 7 | 0.8 | 0.60% | 0.60% |
| Combustion Engines - Automobile (W/$) | 0.00% | 0 | 1 | #NUM! | #NUM! |
| Combustion Engines - Automobile (W/kg) | 0.00% | 0 | 1 | #NUM! | #VALUE! |
| Combustion Engines - Automobile (W/L) | 0.70% | 7 | 0.81 | 0.20% | 0.70% |
| Computed Tomography (CT) (1/mm*sec) | 36.70% | 13 | 0.78 | 4.50% | 1.80% |
| Electric Motors (W/kg) | 6.10% | 4 | 0.88 | 2.90% | 1.40% |
| Electric Motors (W/L) | 0.00% | 5 | 1 | 0.00% | 0.00% |
| Electrical Energy Transmission (W*km/$) | 1.80% | 3 | 1 | 0.60% | 1.10% |
| Electrical Energy Transmission (W*km) | 3.00% | 4 | 0.43 | 4.60% | 6.90% |
| Electrical Energy Transmission - AC (W*km/$) | 0.00% | 0 | 1 | #NUM! | #VALUE! |
| Electrical Energy Transmission - AC (W*km) | 0.00% | 0 | 1 | #NUM! | #VALUE! |
| Electrical Energy Transmission - DC (W*km/$) | 1.80% | 3 | 1 | 0.60% | 1.10% |
| Electrical Energy Transmission - DC (W*km) | 3.00% | 4 | 0.43 | 4.60% | 6.90% |
| Electrical Information Transmission (kB/sec*$) | 52.90% | 2 | 1 | #NUM! | 0.00% |
| Electrical Information Transmission (kB/sec) | 46.00% | 2 | 1 | #NUM! | 0.00% |
| Electrochemical Battery Energy Storage (W*hr/$) | 5.70% | 9 | 0.91 | 0.80% | 0.40% |
| Electrochemical Battery Energy Storage (W*hr/kg) | 7.00% | 14 | 0.95 | 0.50% | 0.10% |
| Electrochemical Battery Energy Storage (W*hr/L) | 9.30% | 13 | 0.84 | 1.20% | 0.40% |
| Electronic Computation (1/sec) | 42.60% | 11 | 0.61 | 8.40% | 14.20% |
| Electronic Computation (1/sec*$) | 49.10% | 18 | 0.93 | 2.10% | 1.10% |
| Electronic Computation (1/sec) | 55.70% | 15 | 0.96 | 1.90% | 1.60% |
| Flywheel Energy Storage (kWhr/kg) | 9.00% | 7 | 0.92 | 1.40% | 1.10% |
| Fuel Cell Energy Production (kW/$) | 14.40% | 5 | 0.99 | 0.90% | 0.60% |
| Genome Sequencing (1/$) | 29.30% | 7 | 0.91 | 3.80% | 3.80% |
| Incandescent Artificial Illumination (1000*lm*hr/$) | 0.00% | 0 | 1 | #NUM! | #VALUE! |
| Integrated Circuit Information Storage (transistors/die) | 37.20% | 14 | 1 | 0.60% | 0.20% |
| Integrated Circuit Processors (transistors/die) | 36.30% | 12 | 0.97 | 1.60% | 0.70% |
| LED Artificial Illumination (lm/$) | 24.40% | 13 | 0.98 | 0.80% | 0.30% |
| LED Artificial Illumination (lm) | 36.20% | 15 | 0.97 | 1.20% | 0.60% |
| Magnet Resonance Imaging (MRI) (1/mm*sec*$) | 47.50% | 6 | 0.88 | 7.40% | 11.30% |
| Magnet Resonance Imaging (MRI) (1/mm*sec) | 45.60% | 6 | 0.85 | 8.30% | 13.50% |
| Magnetic Information Storage (megabits/$) | 39.20% | 17 | 0.88 | 2.70% | 1.80% |

| | | | | | |
|---|---|---|---|---|---|
| Magnetic Information Storage (megabits/cc) | 31.90% | 23 | 0.88 | 2.00% | 0.80% |
| Magnetic Information Storage - Hard Disk (megabits/$) | 39.80% | 15 | 0.85 | 3.30% | 4.40% |
| Magnetic Information Storage - Hard Disk (megabits/cc) | 36.10% | 19 | 0.94 | 1.70% | 0.70% |
| Magnetic Information Storage - Tape (megabits/$) | 36.60% | 9 | 0.85 | 4.80% | 2.20% |
| Magnetic Information Storage - Tape (megabits/cc) | 29.60% | 9 | 0.9 | 3.20% | 0.90% |
| Milling Machines (hp/mm) | 3.90% | 3 | 1 | 0.60% | 4.10% |
| Optical Information Storage (Mb/$) | 34.40% | 4 | 0.7 | 24.00% | 19.20% |
| Optical Information Storage (Mb/cc) | 27.10% | 15 | 0.95 | 1.50% | 0.70% |
| Optical Information Transmission (Kb/sec*km*$) | 65.10% | 13 | 0.93 | 3.00% | 2.10% |
| Optical Information Transmission (Kb/sec) | 68.40% | 12 | 0.9 | 4.00% | 2.90% |
| Photolithography (1/micron*$) | 7.70% | 2 | 1 | #NUM! | 0.00% |
| Photolithography (in2/hr*micron*$) | 24.70% | 2 | 1 | #NUM! | 0.00% |
| Photolithography (in2/hr*micron) | 25.90% | 5 | 0.93 | 4.60% | 4.10% |
| Photolithography (in2/hr*$) | 8.10% | 2 | 1 | #NUM! | 0.00% |
| Solar Photovoltaic Energy Generation (kWhr/$) | 10.80% | 34 | 0.98 | 0.30% | 0.10% |
| Solar Photovoltaic Energy Generation (W/$) | 0.4% | 0.1% | 0.1% | 0.1% | 0.1% |
| Superconductivity (1/K) | 9.50% | 7 | 0.73 | 2.90% | 1.40% |
| Wind Turbine Energy Generation (W/$) | 9.20% | 8 | 0.93 | 1.10% | 0.50% |
| Wireless Information Transmission (bits/sec*m2) | 57.70% | 7 | 0.8 | 9.10% | 6.90% |
| Wireless Information Transmission (bits/sec*Hz) | 17.30% | 16 | 0.84 | 1.80% | 0.80% |
| Wireless Information Transmission (kb/sec) | 50.40% | 15 | 0.86 | 3.70% | 0.90% |

**Table A.4: Statistical Measures for the Non-Dominated Technological Improvement Rates of the 72 Domain-Metrics Pairs Since 1970**

## Appendix B: COM Details for All 28 Domains

This section will explain how each of the patent data sets was located using the COM. The domains are grouped by the method that was used to identify the patent set and starts with the direct COM and ends with more complex adaptations.

### Appendix B.1: Direct COM

There were 14 domains that were located using the COM and required no further reading or development of patent sets. In these cases, the most basic version of the COM worked exactly as it was designed, and no extra patent sets were present that required further testing.

#### Capacitor Energy Storage

Initial Search Terms and MPR scores:

| Search Term | Size of Pre-search | IPC | MPR for IPC | UPC | MPR for UPC |
|---|---|---|---|---|---|
| electric capacitor | 10972 | H01G (CAPACITORS; CAPACITORS, RECTIFIERS, DETECTORS, SWITCHING DEVICES, LIGHT-SENSITIVE OR TEMPERATURE-SENSITIVE DEVICES OF THE ELECTROLYTIC TYPE) | 0.21 | 361 (Electricity: electrical systems and devices) | 0.105 |

From this search, and in particular the titles of the patent classes, it was apparent that H01G and 361 were the two most appropriate patent classes for the TD. The patent set was downloaded and read for relevancy and was chosen to be the final patent set.

(CCL:(361) AND ICL:(H01G)) AND (APD:[1976-1-1 TO 2013-7-1]) AND DOCUMENT_TYPE:United States Issued Patent

N= 5944

Relevancy: 0.84

**Integrated Circuit Processors**

Initial Search Terms and MPR scores:

| Search Term | Size of Pre-search | IPC | MPR for IPC | UPC | MPR for UPC |
|---|---|---|---|---|---|
| Integrated circuits | 67274 | H01L (SEMICONDUCTOR DEVICES; ELECTRIC SOLID STATE DEVICES NOT OTHERWISE PROVIDED FOR | 0.192 | 257 (Active solid-state devices (e.g., transistors, solid-state diodes | 0.21 |
| semiconductors | 152741 | H01L (SEMICONDUCTOR DEVICES; ELECTRIC SOLID STATE DEVICES NOT OTHERWISE PROVIDED FOR | 0.53 | 257 (Active solid-state devices (e.g., transistors, solid-state diodes) | 0.53 |
| transistors | 77792 | H01L (SEMICONDUCTOR DEVICES; ELECTRIC SOLID STATE DEVICES NOT OTHERWISE PROVIDED FOR | 0.26 | 257 (Active solid-state devices (e.g., transistors, solid-state diodes | 0.28 |

From these searches, it was apparent that H01L and 257 were the two most appropriate patent classes for the TD. The patent set was downloaded and read for relevancy and was chosen to be the final patent set.

(CCL:(257) AND ICL:(H01L)) AND (APD:[1976-1-1 TO 2013-7-1]) AND DOCUMENT_TYPE:United States Issued Patent

N= 149491

Relevancy: 0.805

**Milling Machines**

Initial Search Terms and MPR scores:

| Search Term | Size of Pre-search | IPC | MPR for IPC | UPC | MPR for UPC |
|---|---|---|---|---|---|
| milling machine | 607 | B23C (MILLING (broaching B23D; broach-milling in making gears B23F; arrangements for copying or controlling B23Q)) | 0.18 | 409 (Gear cutting, milling, or planing) | 0.16 |

From this search, it was apparent that B23C and 409 were the two most appropriate patent classes for the TD. The patent set was downloaded and read for relevancy and was chosen to be the final patent set.

(CCL:(409) AND ICL:(B23C)) AND (APD:[1976-1-1 TO 2013-7-1]) AND DOCUMENT_TYPE:United States Issued Patent

N= 2315

Relevancy: 0.925

## Optical Information Storage

Initial Search Terms and MPR scores:

| Search Term | Size of Pre-search | IPC | MPR for IPC | UPC | MPR for UPC |
|---|---|---|---|---|---|
| optical disc | 6293 | G11B (INFORMATION STORAGE BASED ON RELATIVE MOVEMENT BETWEEN RECORD CARRIER AND TRANSDUCER (recording measured values in a way that does not require playback through a transducer | 0.38 | 369 (Dynamic information storage or retrieval | 0.37 |
| optical disk | 7659 | G11B (INFORMATION STORAGE BASED ON RELATIVE MOVEMENT BETWEEN RECORD CARRIER AND TRANSDUCER (recording measured values in a way that does not require playback through a transducer | 0.39 | 369 (Dynamic information storage or retrieval | 0.387 |
| optical storage | 5145 | G11B (INFORMATION STORAGE BASED ON RELATIVE MOVEMENT BETWEEN RECORD CARRIER AND TRANSDUCER (recording measured values in a way that does not require playback through a transducer | 0.19 | 369 (Dynamic information storage or retrieval | 0.1857 |
| optical information storage | 1525 | G11B (INFORMATION STORAGE BASED ON RELATIVE MOVEMENT BETWEEN RECORD CARRIER AND TRANSDUCER (recording measured values in a way that does not require playback through a transducer | 0.24 | 369 (Dynamic information storage or retrieval | 0.23 |
| compact disc | 2043 | G11B (INFORMATION STORAGE BASED ON RELATIVE MOVEMENT BETWEEN RECORD CARRIER AND TRANSDUCER (recording measured values in a way that does not require playback through a transducer | 0.15 | 206 (Special receptacle or package | 0.095 |
| compact disk | 1269 | G11B (INFORMATION | 0.19 | 369 (Dynamic | 0.13 |

| | | | | | |
|---|---|---|---|---|---|
| | | STORAGE BASED ON RELATIVE MOVEMENT BETWEEN RECORD CARRIER AND TRANSDUCER (recording measured values in a way that does not require playback through a transducer | | information storage or retrieval | |
| optical data storage | 2223 | G11B (INFORMATION STORAGE BASED ON RELATIVE MOVEMENT BETWEEN RECORD CARRIER AND TRANSDUCER (recording measured values in a way that does not require playback through a transducer | 0.24 | 369 (Dynamic information storage or retrieval | 0.22 |
| optical data | 18881 | G11B (INFORMATION STORAGE BASED ON RELATIVE MOVEMENT BETWEEN RECORD CARRIER AND TRANSDUCER (recording measured values in a way that does not require playback through a transducer | 0.15 | 369 (Dynamic information storage or retrieval | 0.18 |
| memory storage | 21182 | G06F (ELECTRIC DIGITAL DATA PROCESSING (computers in which a part of the computation is effected hydraulically or pneumatically | 0.224 | 711 (Electrical computers and digital processing systems: memory | 0.16 |
| optical memory storage | 546 | G11B (INFORMATION STORAGE BASED ON RELATIVE MOVEMENT BETWEEN RECORD CARRIER AND TRANSDUCER (recording measured values in a way that does not require playback through a transducer | 0.13 | 365 (Static information storage and retrieval) | 0.09 |
| optical memory | 3898 | G11B (INFORMATION STORAGE BASED ON RELATIVE MOVEMENT BETWEEN RECORD CARRIER AND TRANSDUCER (recording measured values in a way that does not require playback through a transducer | 0.11 | 369 (Dynamic information storage or retrieval | 0.11 |
| CD storage | 647 | G11B (INFORMATION STORAGE BASED ON RELATIVE MOVEMENT BETWEEN RECORD CARRIER AND TRANSDUCER (recording measured values in a way that does not require playback through a transducer | 0.16 | 206 (Special receptacle or package | 0.08 |

From these searches, it was apparent that G11B and 369 were the two most appropriate patent classes for the TD. The patent set was downloaded and read for relevancy and was chosen to be the final patent set.

(CCL:(369) AND ICL:(G11B)) AND (APD:[1976-1-1 TO 2013-7-1]) AND DOCUMENT_TYPE:United States Issued Patent

N= 23543

Relevancy: 0.815

**Solar Photovoltaic Energy Generation**

Initial Search Terms and MPR scores:

| Search Term | Size of Pre-search | IPC | MPR for IPC | UPC | MPR for UPC |
|---|---|---|---|---|---|
| solar photovoltaic | 991 | H01L (SEMICONDUCTOR DEVICES; ELECTRIC SOLID STATE DEVICES NOT OTHERWISE PROVIDED FOR) | 0.28 | 136 (Batteries: thermoelectric and photoelectric) | 0.34 |

From this search and the previous work done on the renewable energy case study, it was apparent that H01L and 136 were the two most appropriate patent classes for the TD. The patent set was downloaded and read for relevancy and was chosen to be the final patent set.

(CCL:(136) AND ICL:(H01L)) AND (APD:[1976-1-1 TO 2013-7-1]) AND DOCUMENT_TYPE:United States Issued Patent

N= 5203

Relevancy: 0.85

**Superconductivity**

Initial Search Terms and MPR scores:

| Search Term | Size of Pre-search | IPC | MPR for IPC | UPC | MPR for UPC |
|---|---|---|---|---|---|
| superconductivty | 5068 | H01L (SEMICONDUCTOR DEVICES; ELECTRIC SOLID STATE DEVICES NOT OTHERWISE PROVIDED FOR | 0.164 | 505 (Superconductor technology: apparatus, material, process) | 0.604 |
| high temperature superconductivity | 733 | H01L (SEMICONDUCTOR DEVICES; ELECTRIC SOLID STATE DEVICES NOT OTHERWISE PROVIDED FOR | 0.196 | 505 (Superconductor technology: apparatus, material, process | 0.43 |

From this search and the previous work done on the renewable energy case study, it was apparent that H01L and 505 were the two most appropriate patent classes for the TD. The patent set was downloaded and read for relevancy and was chosen to be the final patent set.

(CCL:(505) AND ICL:(H01L)) AND (APD:[1976-1-1 TO 2013-7-1]) AND DOCUMENT_TYPE:United States Issued Patent

N= 1776

Relevancy: 0.845

While some domains revealed the two primary patent classes within the first search, other TDs contained multiple options for the patent overlap data sets which were tested for relevancy and completeness to decide the best patent overlap set. Each of the TDs in this sub-section revealed many potential patent class overlaps, and ended with only one class-overlap as the final patent set.

### Camera Sensitivity

Initial Search Terms and MPR scores:

| Search Term | Size of Pre-search | IPC | MPR for IPC | UPC | MPR for UPC |
|---|---|---|---|---|---|
| digital photography | 157 | H04N (PICTORIAL COMMUNICATION, e.g. TELEVISION | 0.233 | 348 (Television | 0.245 |
| digital camera | 5285 | H04N (PICTORIAL COMMUNICATION, e.g. TELEVISION | 0.28 | 348 (Television | 0.301 |
| dslr | 3 | G03B (APPARATUS OR ARRANGEMENTS FOR TAKING PHOTOGRAPHS OR FOR PROJECTING OR VIEWING THEM; APPARATUS OR ARRANGEMENTS EMPLOYING ANALOGOUS TECHNIQUES USING WAVES OTHER THAN OPTICAL WAVES; ACCESSORIES THEREFOR | 0.333 | 396 (Photography | 0.333 |
| image sensor | 18861 | is H04N (PICTORIAL COMMUNICATION, e.g. TELEVISION | 0.195 | 348 (Television | 0.19 |
| CCD | 4947 | H04N (PICTORIAL COMMUNICATION, e.g. TELEVISION | 0.21 | 257 (Active solid-state devices (e.g., transistors, solid-state diodes)) | 0.1 |
| CMOS | 8789 | H01L (SEMICONDUCTOR DEVICES; ELECTRIC SOLID STATE DEVICES NOT OTHERWISE PROVIDED FOR | 0.233 | 257 (Active solid-state devices (e.g., transistors, solid-state diodes)) | 0.23 |
| charge couple device | 4263 | H01L (SEMICONDUCTOR DEVICES; ELECTRIC SOLID STATE DEVICES NOT OTHERWISE PROVIDED FOR | 0.11 | 257 (Active solid-state devices (e.g., transistors, solid-state diodes | 0.13 |
| complementary metal oxide semiconductor | 740 | H01L (SEMICONDUCTOR DEVICES; ELECTRIC SOLID STATE DEVICES NOT OTHERWISE PROVIDED FOR | 0.25 | 257 (Active solid-state devices (e.g., transistors, solid-state diodes | 0.24 |
| camera modules | 1909 | H04N (PICTORIAL COMMUNICATION, e.g. TELEVISION | 0.22 | 348 (Television | 0.25 |
| digital photo | 771 | H04N (PICTORIAL COMMUNICATION, e.g. TELEVISION | 0.13 | 250 (Radiant energy | 0.09 |

| | | | | | |
|---|---|---|---|---|---|
| digital photonics | 199 | G01T (MEASUREMENT OF NUCLEAR OR X-RADIATION | 0.07 | 250 (Radiant energy | 0.21 |
| digital video camera | 1286 | H04N (PICTORIAL COMMUNICATION, e.g. TELEVISION | 0.288 | 348 (Television | 0.3 |
| camera sensor | 3576 | H04N (PICTORIAL COMMUNICATION, e.g. TELEVISION | 0.222 | 348 (Television | 0.263 |
| photo sensor | 1787 | H01L (SEMICONDUCTOR DEVICES; ELECTRIC SOLID STATE DEVICES NOT OTHERWISE PROVIDED FOR | 0.11 | 250 (Radiant energy | 0.15 |

The patent set was downloaded and read for relevancy and was chosen to be the final patent set.

(CCL:(257) AND ICL:(H04N)) AND (APD:[1976-1-1 TO 2013-7-1]) AND DOCUMENT_TYPE:United States Issued Patent

N= 1843

Relevancy: 0.855

**Electric Motors**

Initial Search Terms and MPR scores:

| Search Term | Size of Pre-search | IPC | MPR for IPC | UPC | MPR for UPC |
|---|---|---|---|---|---|
| electric motor | 37459 | H02K (DYNAMO-ELECTRIC MACHINES | 0.15 | 310 (Electrical generator or motor structure | 0.12 |
| stator | 20019 | H02K (DYNAMO-ELECTRIC MACHINES | 0.37 | 310 (Electrical generator or motor structure | 0.322 |
| rotor | 44367 | H02K (DYNAMO-ELECTRIC MACHINES | 0.26 | 310 (Electrical generator or motor structure | 0.2 |
| electric machine | 14098 | B23H (WORKING OF METAL BY THE ACTION OF A HIGH CONCENTRATION OF ELECTRIC CURRENT ON A WORKPIECE USING AN ELECTRODE WHICH TAKES THE PLACE OF A TOOL; SUCH WORKING COMBINED WITH OTHER FORMS OF WORKING OF METAL | 0.2 | 310 (Electrical generator or motor structure | 0.14 |
| generator | 591838 | G06F (ELECTRIC DIGITAL DATA PROCESSING | 0.17 | 365 (Static information storage and retrieval | 0.1 |
| electric generator | 62238 | H02P (CONTROL OR REGULATION OF ELECTRIC | 0.075 | 290 (Prime-mover dynamo plants | 0.16 |

| | | | | |
|---|---|---|---|---|
| | | MOTORS, GENERATORS, OR DYNAMO-ELECTRIC CONVERTERS; CONTROLLING TRANSFORMERS, REACTORS OR CHOKE COILS | | | |
| winding currents | 10188 | H02P (CONTROL OR REGULATION OF ELECTRIC MOTORS, GENERATORS, OR DYNAMO-ELECTRIC CONVERTERS; CONTROLLING TRANSFORMERS, REACTORS OR CHOKE COILS | 0.14 | 318 (Electricity: motive power systems | 0.13 |
| brushless motor | 2137 | H02K (DYNAMO-ELECTRIC MACHINES | 0.244 | 318 (Electricity: motive power systems | 0.294 |
| electromagnetic coil | 7087 | H01F (MAGNETS; INDUCTANCES; TRANSFORMERS; SELECTION OF MATERIALS FOR THEIR MAGNETIC PROPERTIES | 0.07 | 335 (Electricity: magnetically operated switches, magnets, and electromagnets | 0.12 |
| electric primary mover | 25 | H02P (CONTROL OR REGULATION OF ELECTRIC MOTORS, GENERATORS, OR DYNAMO-ELECTRIC CONVERTERS; CONTROLLING TRANSFORMERS, REACTORS OR CHOKE COILS | 0.16 | 290 (Prime-mover dynamo plants | 0.123 |
| motor | 152382 | H02P (CONTROL OR REGULATION OF ELECTRIC MOTORS, GENERATORS, OR DYNAMO-ELECTRIC CONVERTERS; CONTROLLING TRANSFORMERS, REACTORS OR CHOKE COILS | 0.296 | 318 (Electricity: motive power systems) | 0.28 |
| rotary motor | 8163 | H02K (DYNAMO-ELECTRIC MACHINES | 0.06 | 310 (Electrical generator or motor structure | 0.06 |
| electric windings | 10795 | H02K (DYNAMO-ELECTRIC MACHINES | 0.178 | 310 (Electrical generator or motor structure | 0.153 |
| mechanical commutator | 319 | H02K (DYNAMO-ELECTRIC MACHINES | 0.196 | 310 (Electrical generator or motor structure | 0.189 |
| electric commutator | 1677 | H02K (DYNAMO-ELECTRIC MACHINES | 0.25 | 310 (Electrical generator or motor structure | 0.26 |
| squirrel cage motor | 238 | H02K (DYNAMO-ELECTRIC MACHINES | 0.23 | 310 (Electrical generator or motor structure | 0.236 |
| wound rotor | 1605 | H02K (DYNAMO-ELECTRIC MACHINES | 0.3456 | 310 (Electrical generator or motor structure | 0.34 |
| permanent magnet motor | 3688 | H02K (DYNAMO-ELECTRIC MACHINES | 0.333 | 310 (Electrical generator or motor structure | 0.312 |
| brushless AC | 115 | H02P (CONTROL OR REGULATION OF ELECTRIC MOTORS, GENERATORS, OR DYNAMO-ELECTRIC CONVERTERS; CONTROLLING TRANSFORMERS, REACTORS OR CHOKE COILS | 0.236 | 318 (Electricity: motive power systems | 0.222 |
| induction motor | 3126 | H02P (CONTROL OR REGULATION OF ELECTRIC MOTORS, GENERATORS, OR DYNAMO-ELECTRIC CONVERTERS; CONTROLLING TRANSFORMERS, REACTORS OR CHOKE COILS | 0.232 | 318 (Electricity: motive power systems | 0.272 |

Several of the most promising patent class overlaps were downloaded and read for relevancy.

| Patent Class Overlap Set | Number of Patents | Relevancy |
|---|---|---|
| (CCL:(290) AND ICL:(H02K)) AND (APD:[1976-1-1 TO 2013-7-1]) AND DOCUMENT_TYPE:United States Issued Patent | 768 | 0.16 |
| (CCL:(318) AND ICL:(H02K)) AND (APD:[1976-1-1 TO 2013-7-1]) AND DOCUMENT_TYPE:United States Issued Patent | 2754 | 0.55 |
| (CCL:(310) AND ICL:(H02K)) AND (APD:[1976-1-1 TO 2013-7-1]) AND DOCUMENT_TYPE:United States Issued Patent | 18575 | 0.855 |

After reading the sets, it was clear that the most relevant and complete patent set for the TD is:

(CCL:(310) AND ICL:(H02K)) AND (APD:[1976-1-1 TO 2013-7-1]) AND DOCUMENT_TYPE:United States Issued Patent

N= 18575

Relevancy: 0.855

**Electrical Energy Transmission**

Initial Search Terms and MPR scores:

| Search Term | Size of Pre-search | IPC | MPR for IPC | UPC | MPR for UPC |
|---|---|---|---|---|---|
| power lines | 26491 | H04M (TELEPHONIC COMMUNICATION (circuits for controlling other apparatus via a telephone cable and not involving telephone switching apparatus | 0.07 | 365 (Static information storage and retrieval | 0.07 |
| power cables | 7822 | H01R (ELECTRICALLY-CONDUCTIVE CONNECTIONS; STRUCTURAL ASSOCIATIONS OF A PLURALITY OF MUTUALLY-INSULATED ELECTRICAL CONNECTING ELEMENTS; COUPLING DEVICES; CURRENT COLLECTORS | 0.06 | 174 (Electricity: conductors and insulators | 0.07 |
| transmission line | 21163 | H01P (WAVEGUIDES; RESONATORS, LINES OR OTHER DEVICES OF THE WAVEGUIDE TYPE (operating at optical frequencies | 0.13 | 333 (Wave transmission lines and networks | 0.12 |

| | | | | | |
|---|---|---|---|---|---|
| electrical transmission | 14146 | B60K (ARRANGEMENT OR MOUNTING OF PROPULSION UNITS OR OF TRANSMISSIONS IN VEHICLES; ARRANGEMENT OR MOUNTING OF PLURAL DIVERSE PRIME-MOVERS; AUXILIARY DRIVES; INSTRUMENTATION OR DASHBOARDS FOR VEHICLES; ARRANGEMENTS IN CONNECTION WITH COOLING, AIR INTAKE, GAS EXHAUST, OR FUEL SUPPLY, OF PROPULSION UNITS, IN VEHICLES | 0.052 | 903 (Hybrid electric vehicles (hevs | 0.106 |
| electrical transmission line | 3294 | H01P (WAVEGUIDES; RESONATORS, LINES OR OTHER DEVICES OF THE WAVEGUIDE TYPE (operating at optical frequencies | 0.088 | 333 (Wave transmission lines and networks | 0.112 |
| losless transmission line | 15 | H01P (WAVEGUIDES; RESONATORS, LINES OR OTHER DEVICES OF THE WAVEGUIDE TYPE (operating at optical frequencies | 0.166 | 333 (Wave transmission lines and networks | 0.267 |
| electrical transmission | 14146 | B60K (ARRANGEMENT OR MOUNTING OF PROPULSION UNITS OR OF TRANSMISSIONS IN VEHICLES; ARRANGEMENT OR MOUNTING OF PLURAL DIVERSE PRIME-MOVERS; AUXILIARY DRIVES; INSTRUMENTATION OR DASHBOARDS FOR VEHICLES; ARRANGEMENTS IN CONNECTION WITH COOLING, AIR INTAKE, GAS EXHAUST, OR FUEL SUPPLY, OF PROPULSION UNITS, IN VEHICLES | 0.052 | 903 (Hybrid electric vehicles (hevs | 0.106 |
| electric power transmission | 3915 | B60K (ARRANGEMENT OR MOUNTING OF PROPULSION UNITS OR OF TRANSMISSIONS IN VEHICLES; ARRANGEMENT OR MOUNTING OF PLURAL DIVERSE PRIME-MOVERS; AUXILIARY DRIVES; INSTRUMENTATION OR DASHBOARDS FOR VEHICLES; ARRANGEMENTS IN CONNECTION WITH COOLING, AIR INTAKE, GAS EXHAUST, OR FUEL SUPPLY, OF PROPULSION UNITS, IN VEHICLES | 0.06 | 180 (Motor vehicles | 0.06 |
| power plant transmission | 144 | B60K (ARRANGEMENT OR MOUNTING OF PROPULSION UNITS OR OF TRANSMISSIONS IN VEHICLES; ARRANGEMENT OR MOUNTING OF PLURAL DIVERSE PRIME-MOVERS; AUXILIARY DRIVES; INSTRUMENTATION OR DASHBOARDS FOR VEHICLES; ARRANGEMENTS IN CONNECTION WITH COOLING, AIR INTAKE, GAS EXHAUST, OR FUEL SUPPLY, OF PROPULSION UNITS, IN VEHICLES | 0.062 | 180 (Motor vehicles | 0.08 |
| power grid | 1895 | H02J (CIRCUIT ARRANGEMENTS OR SYSTEMS FOR SUPPLYING OR DISTRIBUTING ELECTRIC POWER; SYSTEMS FOR STORING ELECTRIC ENERGY (power supply circuits for apparatus for measuring X-radiation, gamma radiation, corpuscular radiation or cosmic radiation | 0.0711 | 290 (Prime-mover dynamo plants | 0.08 |
| electric grid | 3194 | H01L (SEMICONDUCTOR DEVICES; ELECTRIC SOLID STATE DEVICES NOT OTHERWISE PROVIDED FOR | 0.1 | 257 (Active solid-state devices (e.g., transistors, solid-state diodes)) | 0.1 |
| electric substations | 184 | H02H (EMERGENCY PROTECTIVE CIRCUIT ARRANGEMENTS (indicating or signalling | 0.0879 | 361 (Electricity: electrical systems and devices | 0.17 |

| | | | | | |
|---|---|---|---|---|---|
| | | undesired working conditions | | | |
| electric current transmission | 1466 | G01R (MEASURING ELECTRIC VARIABLES; MEASURING MAGNETIC VARIABLES | 0.042 | 324 (Electricity: measuring and testing | 0.054 |
| high voltage transmission | 1703 | H03K (PULSE TECHNIQUE (measuring pulse characteristics | 0.063 | 363 (Electric power conversion systems | 0.0382 |
| HVDC | 63 | H02J (CIRCUIT ARRANGEMENTS OR SYSTEMS FOR SUPPLYING OR DISTRIBUTING ELECTRIC POWER; SYSTEMS FOR STORING ELECTRIC ENERGY (power supply circuits for apparatus for measuring X-radiation, gamma radiation, corpuscular radiation or cosmic radiation | 0.21 | 363 (Electric power conversion systems | 0.25 |
| high voltage direct current | 2075 | H02M (APPARATUS FOR CONVERSION BETWEEN AC AND AC, BETWEEN AC AND DC, OR BETWEEN DC AND DC, AND FOR USE WITH MAINS OR SIMILAR POWER SUPPLY SYSTEMS; CONVERSION OF DC OR AC INPUT POWER INTO SURGE OUTPUT POWER; CONTROL OR REGULATION THEREOF | 0.059 | 363 (Electric power conversion systems) | 0.078 |

Several of the most promising patent class overlaps were downloaded and read for relevancy.

| **Patent Class Overlap Set** | **Number of Patents** | **Relevancy** |
|---|---|---|
| (CCL:(333) AND ICL:(H01R)) AND (APD:[1976-1-1 TO 2013-7-1]) AND DOCUMENT_TYPE:United States Issued Patent | 378 | 0.48 |
| (CCL:(361) AND ICL:(H01P)) AND (APD:[1976-1-1 TO 2013-7-1]) AND DOCUMENT_TYPE:United States Issued Patent | 234 | <1% (only 50 read) |
| (CCL:(363) AND ICL:(H02m)) AND (APD:[1976-1-1 TO 2013-7-1]) AND DOCUMENT_TYPE:United States Issued Patent | 10787 | 0.855 |

After reading the sets, it was clear that the most relevant and complete patent set for the TD is:

(CCL:(363) AND ICL:(H02M)) AND (APD:[1976-1-1 TO 2013-7-1]) AND DOCUMENT_TYPE:United States Issued Patent

N= 10787

Relevancy: 0.855

## Electrical Information Transmission

Initial Search Terms and MPR scores:

| Search Term | Size of Pre-search | IPC | MPR for IPC | UPC | MPR for UPC |
|---|---|---|---|---|---|
| coaxial cable | 4561 | H01R (ELECTRICALLY-CONDUCTIVE CONNECTIONS; STRUCTURAL ASSOCIATIONS OF A PLURALITY OF MUTUALLY-INSULATED ELECTRICAL CONNECTING ELEMENTS; COUPLING DEVICES; CURRENT COLLECTORS | 0.18 | 439 (Electrical connectors | 0.18 |
| single cable | 3590 | G02B (OPTICAL ELEMENTS, SYSTEMS, OR APPARATUS | 0.043 | 174 (Electricity: conductors and insulators | 0.053 |
| shielded cable | 3143 | H01R (ELECTRICALLY-CONDUCTIVE CONNECTIONS; STRUCTURAL ASSOCIATIONS OF A PLURALITY OF MUTUALLY-INSULATED ELECTRICAL CONNECTING ELEMENTS; COUPLING DEVICES; CURRENT COLLECTORS | 0.22 | 439 (Electrical connectors | 0.22 |
| radio frequency cable | 660 | H04B (TRANSMISSION | 0.11 | 455 (Telecommunications | 0.12 |
| telecommunication cable | 822 | G02B (OPTICAL ELEMENTS, SYSTEMS, OR APPARATUS | 0.1234 | 385 (Optical waveguides | 0.124 |
| data transmission | 36050 | H04L (TRANSMISSION OF DIGITAL INFORMATION, e.g. TELEGRAPHIC COMMUNICATION | 0.175 | 370 (Multiplex communications | 0.236 |
| information transmission | 19541 | H04L (TRANSMISSION OF DIGITAL INFORMATION, e.g. TELEGRAPHIC COMMUNICATION | 0.128 | 370 (Multiplex communications | 0.19 |
| telegraph | 161 | H04L (TRANSMISSION OF DIGITAL INFORMATION, e.g. TELEGRAPHIC COMMUNICATION | 0.236 | 178 (Telegraphy | 0.157 |

Several of the most promising patent class overlaps were downloaded and read for relevancy.

| Patent Class Overlap Set | Number of Patents | Relevancy |
|---|---|---|
| (CCL:(333) AND ICL:(H01P)) AND (APD:[1976-1-1 TO 2013-7-1]) AND DOCUMENT_TYPE:United States Issued Patent | 7774 | <1% (only 50 read) |
| (CCL:(439) AND ICL:(H01R)) AND (APD:[1976-1-1 TO 2013-7-1]) AND DOCUMENT_TYPE:United States Issued Patent | 46701 | 0.6675 |

After reading the sets, it was clear that the most relevant and complete patent set for the TD is:

(CCL:(439) AND ICL:(H01R)) AND (APD:[1976-1-1 TO 2013-7-1]) AND DOCUMENT_TYPE:United States Issued Patent

N= 46701

Relevancy: 0.6675

**Electronic Computation**

Initial Search Terms and MPR scores:

| Search Term | Size of Pre-search | IPC | MPR for IPC | UPC | MPR for UPC |
|---|---|---|---|---|---|
| CPU | 12334 | G06F (ELECTRIC DIGITAL DATA PROCESSING (computers in which a part of the computation is effected hydraulically or pneumatically | 0.266 | 710 (Electrical computers and digital data processing systems: input/output) | 0.094 |
| computer | 164410 | G06F (ELECTRIC DIGITAL DATA PROCESSING | 0.36 | 709 (Electrical computers and digital processing systems: multicomputer data transferring | 0.182 |
| central procesing unit | 6650 | G06F (ELECTRIC DIGITAL DATA PROCESSING | 0.2567 | 710 (Electrical computers and digital data processing systems: input/output | 0.08 |
| electronic computer | 11381 | G06F (ELECTRIC DIGITAL DATA PROCESSING | 0.21 | 705 (Data processing: financial, business practice, management, or cost/price determination | 0.083 |
| micro computer | 862 | G06F (ELECTRIC DIGITAL DATA PROCESSING | 0.173 | 710 (Electrical computers and digital data processing systems: input/output | 0.04 |
| micro processor | 724 | G06F (ELECTRIC DIGITAL DATA PROCESSING | 0.21 | 712 (Electrical computers and digital processing systems: processing architectures and instruction processing (e.g., processors)) | 0.09 |
| microprocessor | 18704 | G06F (ELECTRIC DIGITAL DATA PROCESSING | 0.164 | 712 (Electrical computers and digital processing systems: processing architectures and instruction processing (e.g., processors | 0.083 |
| minicomputer | 106 | G06F (ELECTRIC DIGITAL DATA PROCESSING | 0.23 | 370 (Multiplex communications | 0.038 |
| mini computure | 174 | G06F (ELECTRIC DIGITAL DATA PROCESSING | 0.193 | 369 (Dynamic information storage or retrieval | 0.0375 |

| | | | | | |
|---|---|---|---|---|---|
| personal computer | 8352 | G06F (ELECTRIC DIGITAL DATA PROCESSING | 0.222 | 709 (Electrical computers and digital processing systems: multicomputer data transferring | 0.058 |
| electronic mainframe | 42 | G06F (ELECTRIC DIGITAL DATA PROCESSING | 0.262 | 705 (Data processing: financial, business practice, management, or cost/price determination | 0.09 |
| mainframe | 631 | G06F (ELECTRIC DIGITAL DATA PROCESSING | 0.225 | 361 (Electricity: electrical systems and devices | 0.082 |
| computer workstation | 1021 | G06F (ELECTRIC DIGITAL DATA PROCESSING | 0.275 | 709 (Electrical computers and digital processing systems: multicomputer data transferring | 0.11 |

Several of the most promising patent class overlaps were downloaded and read for relevancy.

| Patent Class Overlap Set | Number of Patents | Relevancy |
|---|---|---|
| (CCL:(709) AND ICL:(G06F)) AND APD_YEARMONTHDAY:[19760101 TO 20130701] AND DOCUMENT_TYPE:USB | 57492 | 0.02 |
| ((CCL:(708) AND ICL:(G06F)) AND APD_YEARMONTHDAY:[19760101 TO 20130701] AND DOCUMENT_TYPE:USB) | 10563 | 0.635 |
| ((CCL:(712) AND ICL:(G06F)) AND APD_YEARMONTHDAY:[19760101 TO 20130701] AND DOCUMENT_TYPE:USB) | 14606 | 0.965 |

After reading the sets, it was clear that the most relevant and complete patent set for the TD is:

(CCL:(712) AND ICL:(G06F)) AND (APD:[1976-1-1 TO 2013-7-1]) AND DOCUMENT_TYPE:United States Issued Patent

N= 14606

Relevancy: 0.965

**LED Artificial Illumination**

Initial Search Terms and MPR scores:

| Search Term | Size of Pre-search | IPC | MPR for IPC | UPC | MPR for UPC |
|---|---|---|---|---|---|
| LED | 17094 | F21V (FUNCTIONAL FEATURES OR DETAILS OF LIGHTING DEVICES OR SYSTEMS THEREOF; STRUCTURAL COMBINATIONS OF LIGHTING DEVICES WITH OTHER ARTICLES, NOT OTHERWISE PROVIDED FOR) | 0.13 | 362 (Illumination | 0.17 |
| light emitting diode | 11543 | H01L (SEMICONDUCTOR DEVICES; ELECTRIC SOLID STATE DEVICES NOT OTHERWISE PROVIDED FOR | 0.132 | 257 (Active solid-state devices (e.g., transistors, solid-state diodes)) | 0.1357 |
| semiconductor light | 14672 | H01L (SEMICONDUCTOR DEVICES; ELECTRIC SOLID STATE DEVICES NOT OTHERWISE PROVIDED FOR | 0.2744 | 257 (Active solid-state devices (e.g., transistors, solid-state diodes)) | 0.28 |
| electroluminescent | 4753 | H01L (SEMICONDUCTOR DEVICES; ELECTRIC SOLID STATE DEVICES NOT OTHERWISE PROVIDED FOR | 0.18 | 313 (Electric lamp and discharge devices | 0.3 |
| diode lamp | 1249 | H05B (ELECTRIC HEATING; ELECTRIC LIGHTING NOT OTHERWISE PROVIDED FOR | 0.18 | 315 (Electric lamp and discharge devices: systems | 0.22 |
| solid state light | 3395 | H01L (SEMICONDUCTOR DEVICES; ELECTRIC SOLID STATE DEVICES NOT OTHERWISE PROVIDED FOR | 0.132 | 257 (Active solid-state devices (e.g., transistors, solid-state diodes)) | 0.145 |
| solid state illumination | 468 | H04N (PICTORIAL COMMUNICATION, e.g. TELEVISION | 0.08 | 362 (Illumination | 0.123 |

Several of the most promising patent class overlaps were downloaded and read for relevancy.

| Patent Class Overlap Set | Number of Patents | Relevancy |
|---|---|---|
| (CCL:(362) AND ICL:(F21V)) AND (APD:[1976-1-1 TO 2013-7-1]) AND DOCUMENT_TYPE:United States Issued Patent | 18191 | 0.066 |
| (CCL:(362) AND ICL:(H01L)) AND (APD:[1976-1-1 TO 2013-7-1]) AND DOCUMENT_TYPE:United States Issued Patent | 663 | 0.14 |
| (CCL:(315) AND ICL:(H01L)) AND (APD:[1976-1-1 TO 2013-7-1]) AND DOCUMENT_TYPE:United States Issued Patent | 576 | 0.51 |
| (CCL:(313) AND ICL:(H01L)) AND (APD:[1976-1-1 TO 2013-7-1]) AND DOCUMENT_TYPE:United States Issued Patent | 4043 | 0.85 |

After reading the sets, it was clear that the most relevant and complete patent set for the TD is:

(CCL:(313) AND ICL:(H01L)) AND (APD:[1976-1-1 TO 2013-7-1]) AND DOCUMENT_TYPE:United States Issued Patent

N= 4043

Relevancy: 0.85

The next two cases could be found using the direct COM, as they resulted in a single patent class overlap set. These two data sets are interesting in that they were located by searching only for magnetic information storage. In this case, the search terms for magnetic information storage also led to the identification of the IPC and UPC for integrated circuit information storage.

## Magnetic Information Storage

The initial search terms for magnetic information storage were as follows:

| Search Term | Size of Pre-search | IPC | MPR for IPC | UPC | MPR for UPC |
|---|---|---|---|---|---|
| magnetic tape | 8437 | G11B (INFORMATION STORAGE BASED ON RELATIVE MOVEMENT BETWEEN RECORD CARRIER AND TRANSDUCER (recording measured values in a way that does not require playback through a transducer | 0.37 | 360 (Dynamic magnetic information storage or retrieval | 0.36 |
| tape storage | 2285 | G11B (INFORMATION STORAGE BASED ON RELATIVE MOVEMENT BETWEEN RECORD CARRIER AND TRANSDUCER (recording measured values in a way that does not require playback through a transducer | 0.18 | 360 (Dynamic magnetic information storage or retrieval | 0.15 |
| magnetic information | 7382 | G11B (INFORMATION STORAGE BASED ON RELATIVE MOVEMENT BETWEEN RECORD CARRIER AND TRANSDUCER (recording measured values in a way that does not require playback through a transducer | 0.24 | 360 (Dynamic magnetic information storage or retrieval | 0.2 |
| magnetic disk | 7525 | G11B (INFORMATION STORAGE BASED ON RELATIVE MOVEMENT BETWEEN RECORD | 0.366 | 360 (Dynamic magnetic information storage or retrieval | 0.33 |

| | | | | | |
|---|---|---|---|---|---|
| | | CARRIER AND TRANSDUCER (recording measured values in a way that does not require playback through a transducer | | | |
| magenetic data storage | 1933 | G11B (INFORMATION STORAGE BASED ON RELATIVE MOVEMENT BETWEEN RECORD CARRIER AND TRANSDUCER (recording measured values in a way that does not require playback through a transducer | 0.272 | 360 (Dynamic magnetic information storage or retrieval | 0.25 |
| magnetic data | 9637 | G11B (INFORMATION STORAGE BASED ON RELATIVE MOVEMENT BETWEEN RECORD CARRIER AND TRANSDUCER (recording measured values in a way that does not require playback through a transducer | 0.233 | 360 (Dynamic magnetic information storage or retrieval | 0.234 |
| memory storage | 21182 | G06F (ELECTRIC DIGITAL DATA PROCESSING (computers in which a part of the computation is effected hydraulically or pneumatically | 0.22 | 711 (Electrical computers and digital processing systems: memory) | 0.16 |
| magnetic memory storage | 823 | G11C (STATIC STORES (information storage based on relative movement between record carrier and transducer | 0.2 | 365 (Static information storage and retrieval | 0.21 |
| magnetic memory | 4530 | G11C (STATIC STORES (information storage based on relative movement between record carrier and transducer | 0.22 | 365 (Static information storage and retrieval | 0.23 |
| cassette tape | 4110 | G11B (INFORMATION STORAGE BASED ON RELATIVE MOVEMENT BETWEEN RECORD CARRIER AND TRANSDUCER (recording measured values in a way that does not require playback through a transducer | 0.36 | 360 (Dynamic magnetic information storage or retrieval | 0.323 |
| magnetic cassette | 1969 | G11B (INFORMATION STORAGE BASED ON RELATIVE MOVEMENT BETWEEN RECORD CARRIER AND TRANSDUCER (recording measured values in a way that does not require playback through a transducer | 0.38 | 360 (Dynamic magnetic information storage or retrieval | 0.34 |

The resulting patent classes that were tested had the following results:

| Patent Class Overlap Set | Numer of Patents | Relevancy |
|---|---|---|
| ((CCL:(365) AND ICL:(G11C)) OR ((CCL:(360) AND ICL:(G11B)))) AND APD_YEARMONTHDAY:[19760101 TO 20130701] AND DOCUMENT_TYPE:USB) | 78908 | N/A |
| ((CCL:(365) AND ICL:(G11C)) AND APD_YEARMONTHDAY:[19760101 TO 20130701] AND DOCUMENT_TYPE:USB) | 49018 | N/A |
| ((CCL:(360) AND ICL:(G11B)) AND APD_YEARMONTHDAY:[19760101 TO 20130701] AND DOCUMENT_TYPE:USB) | 33576 | 0.93 |

After reading the patent sets, it was clear that the 360&G11B patent class overlap set was relevant enough to represent magnetic information storage and both of the other two sets were ruled out for relevancy within the first 50 patents read. However, when reading the 365&G11C patent class overlap sets, the reviewer saw many integrated circuit information storage patents, and thus decided to re-read the patent set for relevance to integrated circuit information storage.

The final patent set for Magnetic Information Storage is as follows:

(CCL:(360) AND ICL:(G11B)) AND (APD:[1976-1-1 TO 2013-7-1]) AND DOCUMENT_TYPE:United States Issued Patent

N= 33576

Relevancy: 0.93

**Integrated Circuit Information Storage**

When the 365&G11C patent set was re-read for relevancy for Integrated Circuit memory devices, the relevancy was much higher at 0.81.

| Patent Class Overlap Set | Number of Patents | Relevancy |
|---|---|---|
| ((CCL:(365) AND ICL:(G11C)) AND APD_YEARMONTHDAY:[19760101 TO 20130701] AND DOCUMENT_TYPE:USB) | 49018 | 0.81 |

The final patent set for Integrated Circuit Information Storage is as follows:

(CCL:(365) AND ICL:(G11C)) AND (APD:[1976-1-1 TO 2013-7-1]) AND DOCUMENT_TYPE:United States Issued Patent

N= 49018

Relevancy: 0.81

## Appendix B.2: Combining Multiple Class Overlaps

Other TDs revealed several potential patent class overlaps and the final patent set involved a combination of these patent sets. In some cases, patent class overlaps were combined together, and in other cases, patent classes were removed from an overlap. This sub-section will provide the specific details for each of the TDs that are made up of multiple patent classes.

### Combustion Engines

Initial Search Terms and MPR scores:

| Search Term | Size of Pre- | IPC | MPR for | UPC | MPR for |
|---|---|---|---|---|---|

| | search | | IPC | | UPC |
|---|---|---|---|---|---|
| engine | 124720 | F02B (INTERNAL-COMBUSTION PISTON ENGINES; COMBUSTION ENGINES IN GENERAL | 0.3744 | 123 (Internal-combustion engines | 0.461 |
| combustion engine | 49026 | F02B (INTERNAL-COMBUSTION PISTON ENGINES; COMBUSTION ENGINES IN GENERAL | 0.324 | 123 (Internal-combustion engines | 0.4867 |
| heat engine | 10383 | F02B (INTERNAL-COMBUSTION PISTON ENGINES; COMBUSTION ENGINES IN GENERAL | 0.091 | 123 (Internal-combustion engines | 0.21 |
| internal combustion engine | 41374 | F02B (INTERNAL-COMBUSTION PISTON ENGINES; COMBUSTION ENGINES IN GENERAL | 0.277 | 123 (Internal-combustion engines | 0.475 |
| turbine | 27837 | F02C (GAS-TURBINE PLANTS; AIR INTAKES FOR JET-PROPULSION PLANTS; CONTROLLING FUEL SUPPLY IN AIR-BREATHING JET-PROPULSION PLANTS | 0.391 | 60 (Power plants | 0.212 |
| gas turbine | 13715 | F02C (GAS-TURBINE PLANTS; AIR INTAKES FOR JET-PROPULSION PLANTS; CONTROLLING FUEL SUPPLY IN AIR-BREATHING JET-PROPULSION PLANTS | 0.412 | 60 (Power plants | 0.284 |
| rotary engine | 4001 | F02B (INTERNAL-COMBUSTION PISTON ENGINES; COMBUSTION ENGINES IN GENERAL | 0.206 | 123 (Internal-combustion engines | 0.257 |
| diesel engine | 4246 | F02B (INTERNAL-COMBUSTION PISTON ENGINES; COMBUSTION ENGINES IN GENERAL | 0.142 | 123 (Internal-combustion engines | 0.27 |
| piston | 59496 | F04B (POSITIVE-DISPLACEMENT MACHINES FOR LIQUIDS; PUMPS | 0.14 | 123 (Internal-combustion engines | 0.112 |
| piston engine | 9743 | F02B (INTERNAL-COMBUSTION PISTON ENGINES; COMBUSTION ENGINES IN GENERAL | 0.266 | 123 (Internal-combustion engines | 0.38 |
| two stroke engine | 1522 | F02B (INTERNAL-COMBUSTION PISTON ENGINES; COMBUSTION ENGINES IN GENERAL | 0.312 | 123 (Internal-combustion engines | 0.416 |
| four stroke engine | 688 | F02B (INTERNAL-COMBUSTION PISTON ENGINES; COMBUSTION ENGINES IN GENERAL | 0.3 | 123 (Internal-combustion engines | 0.447 |
| miller cycle | 11 | F01L (CYCLICALLY OPERATING VALVES FOR MACHINES OR ENGINES | 0.273 | 123 (Internal-combustion engines | 0.454 |
| otto cycle | 111 | F02B (INTERNAL-COMBUSTION PISTON ENGINES; COMBUSTION ENGINES IN GENERAL | 0.31 | 123 (Internal-combustion engines | 0.433 |
| wankel engine | 76 | F02B (INTERNAL-COMBUSTION PISTON ENGINES; COMBUSTION ENGINES IN GENERAL | 0.178 | 418 (Rotary expansible chamber devices | 0.2 |
| jet engine | 2754 | F02K (JET-PROPULSION PLANTS | 0.15 | 60 (Power plants | 0.16 |
| turbojet | 470 | F02K (JET-PROPULSION PLANTS | 0.222 | 60 (Power plants | 0.313 |
| turbofan | 529 | F02K (JET-PROPULSION PLANTS | 0.273 | 60 (Power plants | 0.323 |
| ramjet | 205 | F02K (JET-PROPULSION PLANTS | 0.32 | 60 (Power plants | 0.4 |
| rocket | 2585 | F02K (JET-PROPULSION PLANTS | 0.22 | 60 (Power plants | 0.183 |

Several of the most promising patent class overlaps were downloaded and read for relevancy.

| Patent Class Overlap Set | Number of Patents | Relevancy |
|---|---|---|
| (CCL:(123) AND ICL:(F02B)) AND (APD:[1976-1-1 TO 2013-7-1]) AND DOCUMENT_TYPE:United States Issued Patent | 13431 | 0.95 |
| (CCL:(123) AND ICL:(F01L)) AND (APD:[1976-1-1 TO 2013-7-1]) AND DOCUMENT_TYPE:United States Issued Patent | 6719 | 0.985 |
| (CCL:(123) AND (ICL:(F01L) OR ICL:(F02B))) AND (APD:[1976-1-1 TO 2013-7-1]) AND DOCUMENT_TYPE:United States Issued Patent | 19640 | 0.96167072 |

After reading the sets, it was the overlap of several of the patent classes that provided the most relevant and complete patent set for the TD:

(CCL:(123) AND (ICL:(F01L) OR ICL:(F02B))) AND (APD:[1976-1-1 TO 2013-7-1]) AND DOCUMENT_TYPE:United States Issued Patent

N= 19640

Relevancy: 0.962 (this relevancy score was derived from the weighted average of the two overlaps that were combined to create it)

**Computed Tomography (CT)**

Initial search term and MPR score:

| Search Term | Size of Pre-search | IPC | MPR for IPC | UPC | MPR for UPC |
|---|---|---|---|---|---|
| computed tomography | 1360 | A61B (DIAGNOSIS; SURGERY; IDENTIFICATION | 0.232 | 378 (X-ray or gamma ray systems or devices | 0.37 |

There was only one search term necessary for the computed tomography pre-search, as the TD has already been studied in depth in the development of the COM, which can be found in the methodology section.

CCL:(378) AND (ICL:(A61B) OR ICL:(G01N)) AND (APD:[1976-1-1 TO 2013-7-1]) AND (DOCUMENT_TYPE: United States Issued Patent)

N= 7234

Relevancy: 0.88

**Incandescent Artificial Illumination**

Initial Search Terms and MPR scores:

| Search Term | Size of Pre-search | IPC | MPR for IPC | UPC | MPR for UPC |
|---|---|---|---|---|---|
| incandescent light | 1132 | H05B (ELECTRIC HEATING; ELECTRIC LIGHTING NOT OTHERWISE PROVIDED FOR | 0.08 | 362 (Illumination | 0.21 |
| incandescent lamp | 1404 | H01K (ELECTRIC INCANDESCENT LAMPS (details or apparatus or processes for manufacture applicable to both discharge devices and incandescent lamps | 0.174 | 313 (Electric lamp and discharge devices | 0.21 |
| incandescent bulb | 507 | F21V (FUNCTIONAL FEATURES OR DETAILS OF LIGHTING DEVICES OR SYSTEMS THEREOF; STRUCTURAL COMBINATIONS OF LIGHTING DEVICES WITH OTHER ARTICLES, NOT OTHERWISE PROVIDED FOR | 0.12 | 362 (Illumination | 0.23 |
| incandescent globe | 15 | F21V (FUNCTIONAL FEATURES OR DETAILS OF LIGHTING DEVICES OR SYSTEMS THEREOF; STRUCTURAL COMBINATIONS OF LIGHTING DEVICES WITH OTHER ARTICLES, NOT OTHERWISE PROVIDED FOR | 0.133 | 362 (Illumination | 0.3 |

| | | | | | |
|---|---|---|---|---|---|
| incandescent | 2121 | H01K (ELECTRIC INCANDESCENT LAMPS (details or apparatus or processes for manufacture applicable to both discharge devices and incandescent lamps | 0.14 | 313 (Electric lamp and discharge devices | 0.154 |
| filament wire | 1173 | H01J (ELECTRIC DISCHARGE TUBES OR DISCHARGE LAMPS | 0.09 | 313 (Electric lamp and discharge devices | 0.1122 |
| filament | 14793 | D02G (CRIMPING OR CURLING FIBRES, FILAMENTS, YARNS, OR THREADS; YARNS OR THREADS | 0.15 | 428 (Stock material or miscellaneous articles | 0.09 |
| halogen lamp | 943 | H01J (ELECTRIC DISCHARGE TUBES OR DISCHARGE LAMPS | 0.16 | 313 (Electric lamp and discharge devices | 0.24 |

Several of the most promising patent class overlaps were downloaded and read for relevancy.

| Patent Class Overlap Set | Number of Patents | Relevancy |
|---|---|---|
| (CCL:(313) AND ICL:(H01K)) AND (APD:[1976-1-1 TO 2013-7-1]) AND DOCUMENT_TYPE:United States Issued Patent | 952 | 0.58 |
| (CCL:(362) AND ICL:(H01K)) AND (APD:[1976-1-1 TO 2013-7-1]) AND DOCUMENT_TYPE:United States Issued Patent | 186 | 0.395 |
| (CCL:(313) AND ICL:(H01K) AND (NOT (ICL:(H01J1) OR ICL:(F21V)))) AND (APD:[1976-1-1 TO 2013-7-1]) AND DOCUMENT_TYPE:United States Issued Patent | 646 | 0.89 |

After reading the sets, it was an overlap of the 313 and H01K patent classes that provided the most relevant patent set, although it was cleaned with the removal of H01J1 (Details of electrodes, of magnetic control means, of screens, or of the mounting or spacing thereof, common to two or more basic types of discharge tubes or lamps) and F21V because they have very little to do with incandescent lighting. The final set is:

(CCL:(313) AND ICL:(H01K) AND (NOT (ICL:(H01J1) OR ICL:(F21V)))) AND (APD:[1976-1-1 TO 2013-7-1]) AND DOCUMENT_TYPE:United States Issued Patent

N= 646

Relevancy: 0.89

**Magnet Resonance Imaging (MRI)**

Initial Search Terms and MPR scores:

| Search Term | Size of Pre-search | IPC | MPR for IPC | UPC | MPR for UPC |
|---|---|---|---|---|---|
| MRI | 2220 | G01V (GEOPHYSICS; GRAVITATIONAL MEASUREMENTS; DETECTING MASSES OR OBJECTS; TAGS | 0.22 | 324 (Electricity: measuring and testing | 0.29 |
| magnetic resonance imaging | 3610 | G01V (GEOPHYSICS; GRAVITATIONAL MEASUREMENTS; DETECTING MASSES OR OBJECTS; TAGS (means for indicating the location of accidentally buried, e.g. snow-buried, persons | 0.25 | 324 (Electricity: measuring and testing | 0.36 |
| nuclear magnetic resonance imaging | 594 | G01R (MEASURING ELECTRIC VARIABLES; MEASURING MAGNETIC VARIABLES (measuring physical variables of any kind by conversion into electric variables | 0.225 | 324 (Electricity: measuring and testing | 0.41 |
| magnetic resonance tomography | 363 | A61B (DIAGNOSIS; SURGERY; IDENTIFICATION (analysing biological material | 0.14 | 324 (Electricity: measuring and testing | 0.284 |
| medical imaging | 3347 | A61B (DIAGNOSIS; SURGERY; IDENTIFICATION (analysing biological material | 0.2 | 600 (Surgery | 0.193 |
| radiology | 654 | A61B (DIAGNOSIS; SURGERY; IDENTIFICATION (analysing biological materi | 0.093 | 378 (X-ray or gamma ray systems or devices | 0.19 |
| nuclear magnetic resonance | 1568 | G01R (MEASURING ELECTRIC VARIABLES; MEASURING MAGNETIC VARIABLES (measuring physical variables of any kind by conversion into electric variables, | 0.23 | 324 (Electricity: measuring and testing | 0.42 |

Several of the most promising patent class overlaps were downloaded and read for relevancy.

| Patent Class Overlap Set | Number of Patents | Relevancy |
|---|---|---|
| (((CCL:(324) AND ICL:(A61B)) AND APD_YEARMONTHDAY:[19760101 TO 20130701] AND DOCUMENT_TYPE:USB)) | 1464 | <1% (read only 50) |
| (((CCL:(324) AND ICL:(A61B)) OR ((CCL:(600) AND ICL:(G01R)))) AND APD_YEARMONTHDAY:[19760101 TO 20130701] AND DOCUMENT_TYPE:USB) | 1977 | 0.855 |

After reading the sets, it was the overlap of several of the patent classes that provided the most relevant and complete patent set for the TD:

(((CCL:(324) AND ICL:(A61B)) OR ((CCL:(600) AND ICL:(G01R)))) AND APD_YEARMONTHDAY:[19760101 TO 20130701] AND DOCUMENT_TYPE:USB) AND (DOCUMENT_TYPE:USA OR DOCUMENT_TYPE:USB)

N= 1977

Relevancy: 0.855

**Optical Information Transmission**

Initial Search Terms and MPR scores:

| **Search Term** | **Size of Pre-search** | **IPC** | **MPR for IPC** | **UPC** | **MPR for UPC** |
|---|---|---|---|---|---|
| fiberoptic | 501 | A61B (DIAGNOSIS; SURGERY; IDENTIFICATION | 0.143 | 385 (Optical waveguides | 0.16 |
| fiberoptic cable | 97 | G02B (OPTICAL ELEMENTS, SYSTEMS, OR APPARATUS | 0.18 | 385 (Optical waveguides | 0.222 |
| optical data transmission | 2151 | H04B (TRANSMISSION | 0.18 | 398 (Optical communications | 0.26 |
| optical cables | 6408 | G02B (OPTICAL ELEMENTS, SYSTEMS, OR APPARATUS | 0.31 | 385 (Optical waveguides | 0.336 |
| optical fiber cable | 4997 | G02B (OPTICAL ELEMENTS, SYSTEMS, OR APPARATUS | 0.33 | 385 (Optical waveguides | 0.36 |
| fiber optic | 32984 | G02B (OPTICAL ELEMENTS, SYSTEMS, OR APPARATUS | 0.35 | 385 (Optical waveguides | 0.389 |
| optical fiber | 32984 | G02B (OPTICAL ELEMENTS, SYSTEMS, OR APPARATUS | 0.348 | 385 (Optical waveguides | 0.489 |
| fibre-optic cable | 96 | G02B (OPTICAL ELEMENTS, SYSTEMS, OR APPARATUS | 0.266 | 385 (Optical waveguides | 0.26 |

Several of the most promising patent class overlaps were downloaded and read for relevancy.

| Patent Class Overlap Set | Number of Patents | Relevancy |
|---|---|---|
| ((CCL:(398) AND ICL:(H04B)) AND APD_YEARMONTHDAY:[19760101 TO 20130701] AND DOCUMENT_TYPE:USB) | 9153 | 0.92 |
| (CCL:(385) AND ICL:(G02B)) AND (APD:[1976-1-1 TO 2013-7-1]) AND DOCUMENT_TYPE:United States Issued Patent | 30937 | 0.815 |
| ((CCL:(398) AND ICL:(H04B)) OR (CCL:(385) AND ICL:(G02B))) AND APD_YEARMONTHDAY:[19760101 TO 20130701] AND DOCUMENT_TYPE:USB | 36494 | 0.839 |

After reading the sets, it was the overlap of several of the patent classes that provided the most relevant and complete patent set for the TD:

((CCL:(398) AND ICL:(H04B)) OR (CCL:(385) AND ICL:(G02B))) AND APD_YEARMONTHDAY:[19760101 TO 20130701] AND DOCUMENT_TYPE:USB

N= 36494

Relevancy: 0.839 (this relevancy score was derived from the weighted average of the two overlaps that were combined to create it)

### Photolithography

Initial Search Terms and MPR scores:

| Search Term | Size of Pre-search | IPC | MPR for IPC | UPC | MPR for UPC |
|---|---|---|---|---|---|
| optical lithography | 1054 | G03F (PHOTOMECHANICAL PRODUCTION OF TEXTURED OR PATTERNED SURFACES, e.g. FOR PRINTING, FOR PROCESSING OF SEMICONDUCTOR DEVICES) | 0.147 | 355 (Photocopying) | 0.11 |
| photlithography | 3077 | G03F (PHOTOMECHANICAL PRODUCTION OF TEXTURED OR PATTERNED SURFACES, e.g. FOR PRINTING, FOR PROCESSING OF SEMICONDUCTOR DEVICES) | 0.154 | 430 (Radiation imagery chemistry: process, composition, or product thereof) | 0.15 |

Several of the most promising patent class overlaps were downloaded and read for relevancy.

| Patent Class Overlap Set | Number of Patents | Relevancy |
|---|---|---|
| (CCL:(430) AND ICL:(G03F)) AND (APD:[1976-1-1 TO 2013-7-1]) AND DOCUMENT_TYPE:United States Issued Patent | 14785 | 0.865 |
| (CCL:(355) AND ICL:(G03F)) AND (APD:[1976-1-1 TO 2013-7-1]) AND DOCUMENT_TYPE:United States Issued Patent | 1722 | 0.88 |
| ((CCL:(430) OR CCL:(355)) AND ICL:(G03F)) AND (APD:[1976-1-1 TO 2013-7-1]) AND DOCUMENT_TYPE:United States Issued Patent | 15948 | 0.866564791 |

After reading the sets, it was the overlap of several of the patent classes that provided the most relevant and complete patent set for the TD:

((CCL:(430) OR CCL:(355)) AND ICL:(G03F)) AND (APD:[1976-1-1 TO 2013-7-1]) AND DOCUMENT_TYPE:United States Issued Patent

N= 15948

Relevancy: 0.8665 (this relevancy score was derived from the weighted average of the two overlaps that were combined to create it)

### Wireless Information Transmission

Initial Search Terms and MPR scores:

| Search Term | Size of Pre-search | IPC | MPR for IPC | UPC | MPR for UPC |
|---|---|---|---|---|---|
| wireless communication | 23498 | H04B (TRANSMISSION (transmission systems for measured values, control | 0.176 | 455 (Telecommunications | 0.324 |

| | | | | | |
|---|---|---|---|---|---|
| | | or similar signals | | | |
| radio transmission | 7667 | H04B (TRANSMISSION (transmission systems for measured values, control or similar signals | 0.21 | 455 (Telecommunications | 0.18 |
| radio data transmission | 2768 | H04B (TRANSMISSION (transmission systems for measured values, control or similar signals | 0.18 | 370 (Multiplex communications | 0.24 |
| radio wave transmission | 724 | H04B (TRANSMISSION (transmission systems for measured values, control or similar signals | 0.169 | 455 (Telecommunications | 0.21 |
| radio wave | 3258 | H04B (TRANSMISSION (transmission systems for measured values, control or similar signals | 0.124 | 455 (Telecommunications | 0.16 |
| cell phone | 1197 | H04M (TELEPHONIC COMMUNICATION (circuits for controlling other apparatus via a telephone cable and not involving telephone switching apparatus | 0.115 | 455 (Telecommunications | 0.212 |
| cell phone tower | 16 | H04M (TELEPHONIC COMMUNICATION (circuits for controlling other apparatus via a telephone cable and not involving telephone switching apparatus | 0.156 | 455 (Telecommunications | 0.28 |
| radio frequency | 19726 | H04M (TELEPHONIC COMMUNICATION (circuits for controlling other apparatus via a telephone cable and not involving telephone switching apparatus | 0.163 | 455 (Telecommunications | 0.196 |
| RF transmitter | 2201 | H04M (TELEPHONIC COMMUNICATION (circuits for controlling other apparatus via a telephone cable and not involving telephone switching apparatus | 0.188 | 455 (Telecommunications | 0.21 |
| radio emitter | 237 | H03F (AMPLIFIERS | 0.076 | 455 (Telecommunications | 0.14 |
| radio trasmitter | 5070 | H04B (TRANSMISSION | 0.178 | 455 (Telecommunications | 0.219 |
| wireless networking | 14312 | H04L (TRANSMISSION OF DIGITAL INFORMATION, e.g. TELEGRAPHIC COMMUNICATION | 0.126 | 455 (Telecommunications | 0.32 |
| telemetry | 1716 | A61N (ELECTROTHERAPY; MAGNETOTHERAPY; RADIATION | 0.113 | 340 (Communications: electrical | 0.152 |

| | | | | | |
|---|---|---|---|---|---|
| | | THERAPY; ULTRASOUND THERAPY | | | |
| microwave communication | 719 | H04B (TRANSMISSION | 0.099 | 219 (Electric heating | 0.097 |
| electromagnetic information | 1691 | G01S (RADIO DIRECTION-FINDING; RADIO NAVIGATION; DETERMINING DISTANCE OR VELOCITY BY USE OF RADIO WAVES; LOCATING OR PRESENCE-DETECTING BY USE OF THE REFLECTION OR RERADIATION OF RADIO WAVES; ANALOGOUS ARRANGEMENTS USING OTHER WAVES | 0.036 | 340 (Communications: electrical | 0.087 |
| wireless data | 12801 | H04B (TRANSMISSION | 0.14 | 455 (Telecommunications | 0.244 |
| wireless data transfer | 1002 | H04L (TRANSMISSION OF DIGITAL INFORMATION, e.g. TELEGRAPHIC COMMUNICATION | 0.103 | 455 (Telecommunications | 0.208 |

Several of the most promising patent class overlaps were downloaded and read for relevancy.

| **Patent Class Overlap Set** | **Number of Patents** | **Relevancy** |
|---|---|---|
| (CCL:(455) AND ICL:(H04B)) AND (APD:[1976-1-1 TO 2013-7-1]) AND DOCUMENT_TYPE:United States Issued Patent | 37367 | 0.945 |
| ((CCL:(455) AND ICL:(H04L)) AND APD_YEARMONTHDAY:[19760101 TO 20130701] AND DOCUMENT_TYPE:USB) AND (DOCUMENT_TYPE:USA OR DOCUMENT_TYPE:USB) | 9029 | 0.905 |

After reading the sets, it was the overlap of several of the patent classes that provided the most relevant and complete patent set for the TD:

(CCL:(455) AND (ICL:(H04L) OR ICL:(H04B))) AND (APD:[1976-1-1 TO 2013-7-1]) AND DOCUMENT_TYPE:United States Issued Patent

N= 39675

Relevancy: 0.94 (this relevancy score was derived from the weighted average of the two overlaps that were combined to create it)

**Wind Turbines**

The process of selecting the Wind Turbine patent set is described in the methodology section. The final patent set is as follows:

CCL:(378) AND (ICL:(A61B) OR ICL:(G01N)) AND (APD:[1976-1-1 TO 2013-7-1]) AND (DOCUMENT_TYPE: United States Issued Patent)

N= 2498

Relevancy: 0.94

## Appendix B.3: Lower Level Hierarchy Classifications and Keyword Modifications

Finally, there were some TDs that required even more sophisticated uses of the COM. In some cases more precise patent classes were used (5 digit IPCs instead of the usual 4 digit). In other cases, the initial search keywords were supplemented with company names. There were also cases in which multiple domains were found in areas that

There were two cases where the hierarchy level of the patent classes that is generally used in the COM was not sufficient to locate an appropriate set of patents for a TD. These two cases required using more specific patent classes in the IPC and the UPC systems.

**3D-Printing (industrial stereolithography)**

The initial search term for the TD was only 'stereolithography'.

| Search Term | Size of Pre-search | IPC | MPR for IPC | UPC | MPR for UPC |
|---|---|---|---|---|---|
| stereolithography | 234 | B29C (SHAPING OR JOINING OF PLASTICS; SHAPING OF SUBSTANCES IN A PLASTIC STATE, IN GENERAL; AFTER-TREATMENT OF THE SHAPED PRODUCTS, e.g. REPAIRING) | 0.33 | 264 (Plastic and nonmetallic article shaping or treating: processes) | 0.35 |

The results of the COM looked promising, however, the resulting set contained a very large number of patents with very little relevancy.

| Patent Class Overlap Set | Number of Patents | Relevancy |
|---|---|---|
| ((CCL:(264) AND ICL:(B29C)) AND APD_YEARMONTHDAY:[19760101 TO 20130701] AND DOCUMENT_TYPE:USB) | 16721 | 0.085 |

The very high MPRs and the very low relevancy score provided an indication that this case should be explored further. The large disparity in sizes between the keyword search and the resulting patent set is another indicator of a need for emendations to the direct COM.

In order to look further, the patent sets were analyzed and the structure surrounding the IPC and UPCs that were selected was explored and some of the main companies that are associated with SLA 3D Printing were searched (ex: 3D Systems) and their patents were analyzed for the prevailing patent classifications (at the lower hierarchy level). After looking deeper into the patent classes from the prior patent class overlap and the relevant companies' patents, it became apparent that only one sub-class of 264 was very related to the search term - 264/401 (STEREOLITHOGRAPHIC SHAPING FROM LIQUID PRECURSOR). This subclass provided one portion of the issue, however when the 264/401 UPC was overlapped with the B29C IPC, there were still a number of patents that were not related to 3D printing along with a majority that were relevant. In order to only keep the relevant patents, the IPC was narrowed down further to B29C35/08 (Heating, cooling or curing, e.g. crosslinking, vulcanising; Apparatus therefor - by wave energy or particle radiation). Once again the complementary nature of the classification systems (form and function) provides an overlapping patent set for the TD. Finally, due to the very small size of the patent sets, further patents were added by including the keyword stereolithography in the final data set.

((CCL:(264/401) AND ICL:(B29C35/08)) AND (APD:[1976-1-1 TO 2013-7-1]) OR TTL:(stereolithography) AND ((APD:[1976-1-1 TO 2013-7-1]) AND DOCUMENT_TYPE:United States Issued Patent)

N= 251

Relevancy: 0.93

**Flywheel Energy Storage**

The initial search terms for the TD included a number of different options that described mechanical energy storage using flywheels:

| Search Term | Size of Pre-search | IPC | MPR for IPC | UPC | MPR for UPC |
|---|---|---|---|---|---|
| flywheel | 3525 | F16D (COUPLINGS FOR TRANSMITTING ROTATION | 0.105 | 74 (Machine element or mechanism | 0.12 |
| mechanical flywheel | 720 | F16D (COUPLINGS FOR TRANSMITTING | 0.088 | 192 (192 clutches and power-stop control | 0.11 |
| mechanical energy storage | 696 | B60K (ARRANGEMENT OR MOUNTING OF PROPULSION UNITS OR OF TRANSMISSIONS IN VEHICLES; ARRANGEMENT OR MOUNTING OF PLURAL DIVERSE PRIME-MOVERS; AUXILIARY DRIVES; INSTRUMENTATION OR DASHBOARDS FOR VEHICLES; ARRANGEMENTS IN CONNECTION WITH COOLING, AIR INTAKE, GAS EXHAUST, OR FUEL SUPPLY, OF PROPULSION UNITS, IN VEHICLES | 0.034 | 60 (Power plants | 0.0655 |
| kinetic energy storage | 195 | B60K (ARRANGEMENT OR MOUNTING OF PROPULSION UNITS OR OF TRANSMISSIONS IN VEHICLES; ARRANGEMENT OR MOUNTING OF PLURAL DIVERSE PRIME-MOVERS; AUXILIARY DRIVES; INSTRUMENTATION OR DASHBOARDS FOR VEHICLES; ARRANGEMENTS IN CONNECTION WITH COOLING, AIR INTAKE, GAS EXHAUST, OR FUEL SUPPLY, OF PROPULSION UNITS, IN VEHICLES | 0.093 | 180 (Motor vehicles | 0.09 |
| rotational energy | 7700 | F03D (WIND MOTORS | 0.063 | 290 (Prime-mover dynamo plants | 0.068 |
| rotational energy storage | 523 | H02K (DYNAMO-ELECTRIC MACHINES | 0.07 | 290 (Prime-mover dynamo plants | 0.07 |

The result of this showed that International Patent Class F16D was a potentially relevant patent class, and 74 appeared to be the most related US patent class.

| Patent Class Overlap Set | Number of Patents | Relevancy |
|---|---|---|
| ((CCL:(74) AND ICL:(F16D)) AND APD_YEARMONTHDAY:[19760101 TO 20130701] AND DOCUMENT_TYPE:USB) | 2162 | 0.06 |
| ((CCL:(74.572) AND (ICL:(F16F15) OR ICL:(H02K7)) AND APD_YEARMONTHDAY:[19760101 TO 20130701] AND DOCUMENT_TYPE:USB) | 270 | 0.434782609 |

After reading the simple COM overlaps of 74 and F16D, the relevancy scores were very low, therefore a deeper level hierarchy of the US and international patent classification systems was used to gain a more precise set of patents. The next step involved including the overlap of the 74.572 (Power Generating Flywheel) patent class and the F16F15 (Suppression of Vibrations in Systems) and H02K7 (Arrangements for handling mechanical energy structurally associated with the machine). These particular patent classes were chosen by reading through the patents that were found in the initial overlap of 74 and F16D. Even with the overlapping of lower level hierarchy classifications, there were still some non-relevant patents in the set, which greatly impacted the relevancy percentages due to the small size of the patent set. These patents were removed by searching through their titles and abstracts and removing the keywords of patents that were not related to flywheel energy storage. The resulting search term is as follows:

((CCL:(74.572) AND (ICL:(F16DF15) OR ICL:(H02K7)) NOT ((TTL:(engine) OR TTL:(balance) OR TTL:(damp)) OR (ABST:(engine) OR ABST:(balance) OR ABST:(damp))) AND APD_YEARMONTHDAY:[19760101 TO 20130701] AND DOCUMENT_TYPE:USB)

N= 159

Relevancy: 0.70

**Genome Sequencing**

Due to the fact that Genome Sequencing is a rather complex field where IP is important (and thus it is more likely that many different keywords are used when patenting), there were 18 pre-search terms for genome sequencing:

| Search Term | Size of Pre-search | IPC | MPR for IPC | UPC | MPR for UPC |
|---|---|---|---|---|---|
| genome sequencing | 1848 | C12N (MICRO-ORGANISMS OR ENZYMES; COMPOSITIONS THEREOF | 0.323 | 435 (Chemistry: molecular biology and microbiology | 0.44 |
| whole genome sequencing | 38 | C07H (SUGARS; DERIVATIVES THEREOF; NUCLEOSIDES; NUCLEOTIDES; NUCLEIC ACIDS | 0.25 | 435 (Chemistry: molecular biology and microbiology | 0.434 |
| complete genome sequencing | 76 | C12N (MICRO-ORGANISMS OR ENZYMES; COMPOSITIONS THEREOF | 0.25 | 435 (Chemistry: molecular biology and microbiology | 0.38 |
| entire genome sequencing | 63 | C12Q (MEASURING OR TESTING PROCESSES INVOLVING ENZYMES OR MICRO-ORGANISMS | 0.27 | 435 (Chemistry: molecular biology and microbiology | 0.43 |
| human genome sequencing | 506 | C12N (MICRO-ORGANISMS OR ENZYMES; COMPOSITIONS THEREOF | 0.33 | 435 (Chemistry: molecular biology and microbiology | 0.422 |
| DNA sequencing | 7973 | C12N (MICRO-ORGANISMS OR ENZYMES; COMPOSITIONS THEREOF | 0.344 | 435 (Chemistry: molecular biology and microbiology | 0.465 |
| DNA assay | 1201 | C12Q (MEASURING OR TESTING PROCESSES INVOLVING ENZYMES OR MICRO-ORGANISMS | 0.262 | 435 (Chemistry: molecular biology and microbiology | 0.45 |
| DNA | 16703 | C12N (MICRO-ORGANISMS OR ENZYMES; COMPOSITIONS THEREOF (biocides, pest repellants or attractants, or plant growth regulators containing micro-organisms, viruses, microbial fungi, enzymes, fermentates, or substances produced by, or extracted from, micro-organisms or animal material | 0.366 | 435 (Chemistry: molecular biology and microbiology | 0.4744 |
| DNA analysis | 718 | C12Q (MEASURING OR TESTING PROCESSES INVOLVING ENZYMES OR MICRO-ORGANISMS | 0.31 | 435 (Chemistry: molecular biology and microbiology | 0.4 |
| SNP genotyping | 41 | C12Q (MEASURING OR TESTING PROCESSES INVOLVING ENZYMES OR MICRO-ORGANISMS | 0.42 | 435 (Chemistry: molecular biology and microbiology | 0.451 |
| DNA genotyping | 132 | C12Q (MEASURING OR TESTING PROCESSES INVOLVING ENZYMES OR MICRO-ORGANISMS | 0.34 | 435 (Chemistry: molecular biology and microbiology | 0.45 |
| genotyping | 746 | C12Q (MEASURING OR TESTING PROCESSES INVOLVING ENZYMES OR MICRO-ORGANISMS | 0.27 | 435 (Chemistry: molecular biology and microbiology | 0.386 |
| assay | 11265 | G01N (INVESTIGATING OR ANALYSING MATERIALS BY DETERMINING THEIR CHEMICAL OR PHYSICAL PROPERTIES (separating components of materials in general | 0.277 | 435 (Chemistry: molecular biology and microbiology | 0.401 |

| | | | | | |
|---|---|---|---|---|---|
| DNA profiling | 123 | C07H (SUGARS; DERIVATIVES THEREOF; NUCLEOSIDES; NUCLEOTIDES; NUCLEIC ACIDS (derivatives of aldonic or saccharic acids | 0.17 | 435 (Chemistry: molecular biology and microbiology | 0.37 |
| genome | 3349 | C12N (MICRO-ORGANISMS OR ENZYMES; COMPOSITIONS THEREOF (biocides, pest repellants or attractants, or plant growth regulators containing micro-organisms, viruses, microbial fungi, enzymes, fermentates, or substances produced by, or extracted from, micro-organisms or animal material | 0.32 | 435 (Chemistry: molecular biology and microbiology | 0.437 |
| high throughput sequencing | 175 | C12Q (MEASURING OR TESTING PROCESSES INVOLVING ENZYMES OR MICRO-ORGANISMS | 0.178 | 435 (Chemistry: molecular biology and microbiology | 0.25 |
| genetic analysis | 274 | C12Q (MEASURING OR TESTING PROCESSES INVOLVING ENZYMES OR MICRO-ORGANISMS | 0.27 | 435 (Chemistry: molecular biology and microbiology | 0.34 |
| genetic | 7903 | C12N (MICRO-ORGANISMS OR ENZYMES; COMPOSITIONS THEREOF (biocides, pest repellants or attractants, or plant growth regulators containing micro-organisms, viruses, microbial fungi, enzymes, fermentates, or substances produced by, or extracted from, micro-organisms or animal material | 0.36 | 435 (Chemistry: molecular biology and microbiology) | 0.41 |

The result of this showed that clearly the US patent class 435 was the most related UPC, and that the IPC could be a number of options including C12N, G07H, and G01N. All of the IPCs were tested for relevancy and none of the direct COM overlaps resulted in a highly relevant set.

| **Patent Class Overlap Set** | **Number of Patents** | **Relevancy** |
|---|---|---|
| (CCL:(435) AND ICL:(C12N)) AND (APD:[1976-1-1 TO 2013-7-1]) AND DOCUMENT_TYPE:United States Issued Patent | *56588* | *0.12* |
| ((CCL:(435) AND ICL:(C07H)) AND APD_YEARMONTHDAY:[19760101 TO 20130701] AND DOCUMENT_TYPE:USB) AND (DOCUMENT_TYPE:USA OR DOCUMENT_TYPE:USB) | *20535* | *0.26* |
| (CCL:(435) AND ICL:(G01N)) AND (APD:[1976-1-1 TO 2013-7-1]) AND DOCUMENT_TYPE:United States Issued Patent | *22324* | *0.22* |
| ((CCL:(435/6.11)) AND ICL:(C12Q)) AND (APD:[1976-1-1 TO 2013-7-1]) AND DOCUMENT_TYPE:United States Issued Patent | *2458* | 0.755 |
| ((CCL:(435/6.12)) AND ICL:(C12Q)) AND (APD:[1976-1-1 TO 2013-7-1]) AND | *2974* | 0.735 |

DOCUMENT_TYPE:United States Issued Patent

The next step was to look closer into the lower level hierarchy patent classification codes by searching for patents from companies that were known to be working in this space:

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

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

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

N= 4861

Relevancy: 0.74

**Aircraft Transport**

For this TD, 8 search terms were used to located the international and US patent classes of interest:

| Search Term | Size of Pre-search | IPC | MPR for IPC | UPC | MPR for UPC |
|---|---|---|---|---|---|
| airplane | 2091 | B64C (AEROPLANES; HELICOPTERS (air-cushion vehicles B60V)) | 0.104 | 244 (Aeronautics and astronautics) | 0.175 |
| aeroplane | 84 | A63H (TOYS, e.g. TOPS, DOLLS, HOOPS, BUILDING BLOCKS) | 0.053 | 239 (Fluid sprinkling, spraying, and diffusing) | 0.0119 |
| aircraft | 19694 | B64D (EQUIPMENT FOR FITTING IN OR TO AIRCRAFT; FLYING SUITS; PARACHUTES; ARRANGEMENTS OR MOUNTING OF POWER PLANTS OR PROPULSION TRANSMISSIONS) | **0.237** | 244 (Aeronautics and astronautics) | **0.32** |
| jet plane | 891 | B41J (TYPEWRITERS; SELECTIVE PRINTING MECHANISMS, i.e. MECHANISMS PRINTING OTHERWISE THAN FROM A FORME; CORRECTION OF TYPOGRAPHICAL ERRORS | 0.1 | 347 (Incremental printing of symbolic information) | 0.125 |
| jet | 32182 | B41J (TYPEWRITERS; SELECTIVE PRINTING MECHANISMS, i.e. MECHANISMS PRINTING OTHERWISE THAN FROM A FORME; CORRECTION OF TYPOGRAPHICAL ERRORS | 0.2 | 347 (Incremental printing of symbolic information) | 0.28 |
| plane | 108823 | G02B (OPTICAL ELEMENTS, SYSTEMS, OR APPARATUS | 0.07 | 343 (Communications: radio wave antennas) | 0.08 |
| fixed wing | 1088 | B64C (AEROPLANES; HELICOPTERS (air-cushion vehicles B60V)) | 0.134 | 244 (Aeronautics and astronautics) | 0.16 |
| fixed wing aircraft | 302 | B64C (AEROPLANES; HELICOPTERS (air-cushion vehicles B60V)) | 0.32 | 244 (Aeronautics and astronautics) | 0.39 |

The result of this showed that UPC 244 was likely the most related US patent class and that either B64C or B64D were the most related IPCs.

| Patent Class Overlap Set | Number of Patents | Relevancy |
|---|---|---|
| (CCL:(244) AND ICL:(B64D)) AND (APD:[1976-1-1 TO 2013-7-1]) AND DOCUMENT_TYPE:United States Issued Patent | 3962 | 0.665 |
| (CCL:(244) AND ICL:(B64C)) AND (APD:[1976-1-1 TO 2013-7-1]) AND DOCUMENT_TYPE:United States Issued Patent | 6225 | 0.8 |

Ultimately the combination of the both of those overlaps was the final patent set for the Aircraft Transportation TD. Along with combining the two previous sets, a number of the non-relevant patents (related to parachutes and canopies) were removed by keyword selection in the title and abstract. The final search term is as follows:

(CCL:(244) AND (ICL:(B64D) OR ICL:(b64c))) AND (NOT ttl:(canopy))AND (NOT ttl:(parachute)) AND (NOT ttl:(helicopter)) AND (APD:[1976-1-1 TO 2013-7-1]) AND DOCUMENT_TYPE:United States Issued Patent

N= 8946

Relevancy: 0.785

**Batteries/Fuel Cells**

Another special case involved two TDs who returned the same patent class overlap sets. Recall from the methodology section that the electrochemical battery energy storage TD involved a patent class overlap set with a set of search terms removed. Predictably, when the fuel cell TD was being searched, the same IPC and UPC were selected as the most relevant patent classes. The 2 TDs were then filtered within the patent class by the keyword terms so that the fuel cell patent set included all of the patents that included the terms 'fuel cell' in the title or abstract, and the battery patents were the patents that did NOT include that term. This example illustrates the possibility of two TDs comprising one patent class overlap and how that can be dealt with through keyword splicing.

**Electrochemical Battery Energy Storage**

As this TD had been studied previously in the renewable energy study, there was only one search term:

| Search Term | Size of Pre-search | IPC | MPR for IPC | UPC | MPR for UPC |
|---|---|---|---|---|---|
| electrochemical cell | 5331 | H01M (PROCESSES OR MEANS, e.g. BATTERIES, FOR THE DIRECT CONVERSION OF CHEMICAL ENERGY INTO ELECTRICAL | 0.41 | 429 (Chemistry: electrical current producing apparatus, product, and process | 0.37 |

The resulting patent class overlap showed marginal reliability

| Patent Class Overlap Set | Number of Patents | Relevancy |
|---|---|---|
| ((CCL:(429) AND ICL:(H01M)) AND APD_YEARMONTHDAY:[19760101 TO 20130701] AND DOCUMENT_TYPE:USB) | 22115 | 0.62 |

However, this data set could be improved by removing the fuel cell patents, and the resulting final patent set is as follows:

((CCL:(429) AND ICL:(H01M)) NOT (TTL:(Fuel Cell))) AND (APD:[1976-1-1 TO 2013-7-1]) AND (DOCUMENT_TYPE:United States Issued Patent)

N= 16122

Relevancy: 0.83

**Fuel Cell Energy Production**

The initial search terms for fuel cells were as follows:

| Search Term | Size of Pre-search | IPC | MPR for IPC | UPC | MPR for UPC |
|---|---|---|---|---|---|
| Fuel cells | 10668 | H01M (PROCESSES OR MEANS, e.g. BATTERIES, | 0.477 | 429 (Chemistry: electrical current | 0.483 |

| | | | | | |
|---|---|---|---|---|---|
| | | FOR THE DIRECT CONVERSION OF CHEMICAL ENERGY INTO ELECTRICAL | | producing apparatus, product, and process | |
| hydrogren fuel cell | 1833 | H01M (PROCESSES OR MEANS, e.g. BATTERIES, FOR THE DIRECT CONVERSION OF CHEMICAL ENERGY INTO ELECTRICAL ENERGY | 0.362 | 429 (Chemistry: electrical current producing apparatus, product, and process | 0.38 |
| methanl fuel cell | 429 | H01M (PROCESSES OR MEANS, e.g. BATTERIES, FOR THE DIRECT CONVERSION OF CHEMICAL ENERGY INTO ELECTRICAL ENERGY | 0.3987 | 429 (Chemistry: electrical current producing apparatus, product, and process | 0.396 |
| anode cathode | 16505 | H01M (PROCESSES OR MEANS, e.g. BATTERIES, FOR THE DIRECT CONVERSION OF CHEMICAL ENERGY INTO ELECTRICAL ENERGY | 0.16 | 204 (Chemistry: electrical and wave energy) | 0.14 |
| ion exchange | 6521 | B01D (SEPARATION (separating solids from solids by wet methods | 0.094 | 210 (Liquid purification or separation | 0.134 |
| proton exchange membrane | 346 | H01M (PROCESSES OR MEANS, e.g. BATTERIES, FOR THE DIRECT CONVERSION OF CHEMICAL ENERGY INTO ELECTRICAL ENERGY | 0.35 | 429 (Chemistry: electrical current producing apparatus, product, and process | 0.38 |
| phosphoric acid fuel cell | 84 | H01M (PROCESSES OR MEANS, e.g. BATTERIES, FOR THE DIRECT CONVERSION OF CHEMICAL ENERGY INTO ELECTRICAL ENERGY | 0.43 | 429 (Chemistry: electrical current producing apparatus, product, and process | 0.453 |
| solid oxide fuel cells | 815 | H01M (PROCESSES OR MEANS, e.g. BATTERIES, FOR THE DIRECT CONVERSION OF CHEMICAL ENERGY INTO ELECTRICAL ENERGY | 0.434 | 429 (Chemistry: electrical current producing apparatus, product, and process | 0.44 |
| molten carbonate fuel cells | 190 | H01M (PROCESSES OR MEANS, e.g. BATTERIES, FOR THE DIRECT CONVERSION OF CHEMICAL ENERGY INTO ELECTRICAL ENERGY | 0.434 | 429 (Chemistry: electrical current producing apparatus, product, and process | 0.452 |
| alkaline fuel cell | 122 | H01M (PROCESSES OR MEANS, e.g. BATTERIES, FOR THE DIRECT CONVERSION OF CHEMICAL ENERGY INTO ELECTRICAL | 0.36 | 429 (Chemistry: electrical current producing apparatus, product, and process | 0.37 |

| | | | | | |
|---|---|---|---|---|---|
| | | ENERGY | | | |

Once it had been established that the search terms did not add any new significant patent classes, it was clear that the fuel cell patent set is the set of patents that had been removed from the battery patent class using the keywords.

(CCL:(429) AND ICL:(H01M)) AND (TTL:(fuel cell) OR ABST:(fuel cell)) AND (APD:[1976-1-1 TO 2013-7-1]) AND DOCUMENT_TYPE:United States Issued Patent

N= 7368

Relevancy: 0.97

# Appendix C: All Domain Metric Pairs tests

A total of 44 DPMs were applied to each of the 28 patent sets and while many of them were discussed in section 3.5, others are listed below. Additionally, table 33 shown below contains the correlation values for all 44 DPMS is shown are shown below.

*Hypothesis 1: Effort*

The following DPMs are relevant to hypothesis 1.

Yearly Average Patenting Rate

This is the average yearly number of patents within a technological domain. In this research, this includes patents that were published between January 1st, 1976 and July 1st, 2013. This measure is calculated using Equation C.1 where SPC is the simple patent count and $\Delta t$ date range in years, in most cases $\Delta t$=37 (2013-1976).

$$\frac{SPC}{\Delta t} \qquad \text{(Equation C.1)}$$

*Hypothesis 2: Importance*

The following DPMs are relevant to hypothesis 2.

Total Number of Forward Citations (duplicates removed)

This is the average number of Forward citations for the patents in a technological domain with the duplicates removed. This measure is calculated using Equation C.2 where SPC is the simple patent count, and $FC_i$ is the number of Forward citations for patent $i$ and $\cup$ is the union of two sets and COUNT() counts the number of elements in a set. Please note that

$\bigcup_{i=1}^{SPC} FC_i$ differs from $\frac{\sum_{i=1}^{SPC}\sum_{j=1}^{FC_i} 1}{SPC}$ in that the duplicates have been removed.

$$COUNT(\bigcup_{i=1}^{SPC} FC_i) \qquad \text{(Equation C.2)}$$

Total Number of U.S. Patent Classes in Citing Patents (duplicates removed)

This is the total number of US patent classes to which the forward citations belong for the patents in a technological domain. This measure is calculated using Equation C.3 where SPC is the simple patent count, $FC_i$ is the number of forward citations for patent *i*, $UPC_{ij}$ is the **set** of US Patent classification codes for each forward citation *j* of patent *i*, and $\bigcup$ is the union of two sets and COUNT() counts the number of elements in a set.

$$COUNT(\bigcup_{i=1}^{SPC}(\bigcup_{j=1}^{FC_i} UPC_{ij}) \qquad \text{(Equation C.3)}$$

Total Number of International Patent Classes in Citing Patents (duplicates removed)

This is the total number of international patent classes to which the forward citations belong for the patents in a technological domain. This measure is calculated using Equation C.4 where SPC is the simple patent count, $FC_i$ is the number of forward citations for patent *i*, $IPC_{ij}$ is the **set** of International Patent classification codes for each forward citation $j$ of patent $i$, and

$\bigcup$ is the union of two sets and COUNT() counts the number of elements in a set.

$$COUNT(\bigcup_{i=1}^{SPC}(\bigcup_{j=1}^{FC_i} IPC_{ij}) \qquad \text{(Equation C.4)}$$

Average Number of U.S. Patent Classes in the Forward Citations Per Patent

This is the average number of US patent classes to which the forward citations belong for each patent in a technological domain. This measure is calculated using Equation C.5 where SPC is the simple patent count, $FC_i$ is the number of forward citations for patent $i$, $UPC_{ij}$ is the **set** of US Patent classification codes for each forward citation $j$ of patent $i$.

$$\frac{\sum_{i=1}^{SPC}\sum_{j=1}^{FC_i} UPC_{ij}}{SPC} \qquad \text{(Equation C.5)}$$

Average Number of International Patent Classes in the Forward Citations Per Patent

This is the average number of International patent classes to which the forward citations belong for each patent in a technological domain. This measure is calculated using Equation C.6 where SPC is the simple patent count, $FC_i$ is the number of forward citations for patent $i$, $IPC_{ij}$ is the **set** of International Patent classification codes for each forward citation $j$ of patent $i$.

$$\frac{\sum_{i=1}^{SPC}\sum_{j=1}^{FC_i} IPC_{ij}}{SPC} \qquad \text{(Equation C.6)}$$

Ratio of Patents with more than 20 Forward Citations

This is the ratio of patents in a technological domain that have received more than 20 citations. This measure is calculated using Equation C.7 where SPC is the simple patent count, $FC_i$ is the number of Forward citations for patent *i*, and the function IF(*arg*) only counts the values if the argument is satisfied. In this situation, $IF(FC_i > 20)$ will only be counted if patent *i* has more than 20 forward citations.

$$\frac{\sum_{i=1}^{SPC} IF(FC_i > 20)}{SPC} \quad \text{(Equation C.7)}$$

*Hypothesis 3: Science*

All of the DPMs that were used to test hypothesis 3 were discussed in section 3.5.3 of the methodology chapter of the thesis.

*Hypothesis 4: Recency*

The following DPMs are relevant to hypothesis 4.

Average Application Year

This is the average year of application for the patents within a technological domain. In this research, this includes patents that were published between January 1st, 1976 and July 1st, 2013. This measure is calculated using Equation C.8 where SPC is the simple patent count and $t_{i_{ay}}$ is the application year of patent *i*.

$$\frac{\sum_{i=1}^{SPC} t_{i_{app}}}{SPC}$$ (Equation C.8)

Average Processing Time per Patent

This is the average time it takes from when a patent is applied for until it is published for the patents within a technological domain. This measure is calculated using Equation C.9 where SPC is the simple patent count, $t_{i_{app}}$ is the application year and $t_{i_{pub}}$ is the publication year of patent *i*.

$$\frac{\sum_{i=1}^{SPC} t_{i_{pub}} - t_{i_{app}}}{SPC}$$ (Equation C.9)

Average number of Forward Citations within 5 years of publication per patent

This is the average number of forward citations that each patent received within 5 years of publication for patents in a technological domain. This measure is calculated using Equation C.10 where SPC is the simple patent count, $FC_i$ is the number of Forward citations for patent *i*, $t_{i_{pub}}$ is the publication year of patent *i*, $t_{ij_{pub}}$ is the publication date of forward citation *j* of patent *i*, and the function IF(*arg*) only counts the values if the argument is satisfied.

$$\sum_{i=1}^{SPC}\sum_{j=1}^{FC_i} IF(t_{ij_{pub}} - t_{i_{pub}} \leq 5)$$ (Equation C.10)

Patents in the 1970s

This is the number of patents published during the 1970s within a technological domain. In this research, this includes patents that were published between January 1st, 1976 and July 1st,

2013. This measure is calculated using Equation C.11 where t is the date, and $P_t$ is the count of the set of patents issued on that particular date.

$$\sum_{t=1/1/1976}^{1/1/1980} P_t \qquad \text{(Equation C.11)}$$

<u>Patents in the 1980s</u>

This is the number of patents published during the 1980s within a technological domain. In this research, these include patents that were published between January 1st, 1976 and July 1st, 2013. This measure is calculated using Equation C.12 where t is the date, and $P_t$ is the count of the set of patents issued on that particular date.

$$\sum_{t=1/1/1980}^{1/1/1990} P_t \qquad \text{(Equation C.12)}$$

<u>Patents in the 1990s</u>

This is the number of patents published during the 1990s within a technological domain. In this research, this includes patents that were published between January 1st, 1976 and July 1st, 2013. This measure is calculated using Equation C.13 where t is the date, and $P_t$ is the count of the set of patents issued on that particular date.

$$\sum_{t=1/1/1990}^{1/1/2000} P_t \qquad \text{(Equation C.13)}$$

Patents since the year 2000

This is the number of patents published since the year 2000 within a technological domain. In this research, this includes patents that were published between January 1st, 1976 and July 1st, 2013. This measure is calculated using Equation C.14 where t is the date, and $P_t$ is the count of the set of patents issued on that particular date.

$$\sum_{t=1/1/2000}^{7/1/2013} P_t \qquad \text{(Equation C.14)}$$

Average Range of Backward Citation Publication Years

This is the average number of years between when a patents oldest backward citation was published and its most recent backward citation was published within a technological domain. This measure is calculated using Equation C.15 where SPC is the simple patent count, $BC_i$ is the number of backward citations for patent $i$, $t_{ji_{pub}}$ is the year of publication of backward citation $j$ of patent $i$ and the function $MAX_i$ finds the maximum value for a certain patent $i$ and the function $MIN_i$ finds the minimum value for a certain patent $i$. Please note that $\sum_{i=1}^{SPC}\sum_{j=1}^{BC_i} 1$ is the just the sum of the total count of backward citations for all of the patents in the TD (without duplicates removed).

$$\frac{\sum_{i=1}^{SPC}\sum_{j=1}^{BC_i} MAX_i(t_{ji_{pub}}) - MIN_i(t_{ji_{pub}})}{\sum_{i=1}^{SPC}\sum_{j=1}^{BC_i} 1} \qquad \text{(Equation C.15)}$$

Average Date of Backward Citation Publication

This is the average date of publication of the backward citations for patents within a technological domain. This measure is calculated using Equation C.16 where SPC is the simple patent count, $BC_i$ **is the number of backward citations for patent *i*,** $t_{ji_{pub}}$ is the year of publication of backward citation *j* of patent *i*. Please note that $\sum_{i=1}^{SPC}\sum_{j=1}^{BC_i} 1$ is the just the sum of the total count of backward citations for all of the patents in the TD (without duplicates removed).

$$\frac{\sum_{i=1}^{SPC}\sum_{j=1}^{BC_i} t_{ji_{pub}}}{\sum_{i=1}^{SPC}\sum_{j=1}^{BC_i} 1} \qquad \text{(Equation C.16)}$$

Average Date of Forward Citation Publication

This is the average date of the patents within a technological domain's Forward citations. This measure is calculated using Equation C.17 where SPC is the simple patent count, $FC_i$ is the number of Forward citations for patent *i*, $t_{ji_{pub}}$ is the year of publication of Forward citation *j* of patent *i*. Please note that $\sum_{i=1}^{SPC}\sum_{j=1}^{FC_i} 1$ is the just the sum of the total count of Forward citations for all of the patents in the TD (without duplicates removed).

$$\frac{\sum_{i=1}^{SPC}\sum_{j=1}^{FC_i} t_{ji_{pub}}}{\sum_{i=1}^{SPC}\sum_{j=1}^{FC_i} 1} \qquad \text{(Equation C.17)}$$

Average Range of Forward Citation Publication Years

This is the average number of years between when a patents oldest Forward citation was published and its most recent Forward citation was published within a technological domain. This measure is calculated using Equation C.18 where SPC is the simple patent count, $BC_i$ is the number of Forward citations for patent *i*, $t_{ji_{pub}}$ is the year of publication of Forward citation *j* of patent *i* and the function $MAX_i$ finds the maximum value for a certain patent *i* and the function $MIN_i$ finds the minimum value for a certain patent *i*. Please note that $\sum_{i=1}^{SPC}\sum_{j=1}^{BC_i}1$ is the just the sum of the total count of Forward citations for all of the patents in the TD (without duplicates removed).

$$\frac{\sum_{i=1}^{SPC}\sum_{j=1}^{BC_i} MAX_i(t_{ji_{pub}}) - MIN_i(t_{ji_{pub}})}{\sum_{i=1}^{SPC}\sum_{j=1}^{BC_i}1} \qquad \text{(Equation C.18)}$$

*Hypothesis 5: Breadth of Knowledge*

The following DPMs are relevant to hypothesis 5.

Average Number of Backward Citations per Patent

This is the average number of backward citations for the patents in a technological domain. This measure is calculated using Equation C.19 where SPC is the simple patent count, and $BC_i$ is the number of backward citations for patent *i*. Please note that $\sum_{i=1}^{SPC}\sum_{j=1}^{BC_i}1$ is the just the sum of the total count of backward citations for all of the patents in the TD (without duplicates removed).

$$\frac{\sum_{i=1}^{SPC}\sum_{j=1}^{BC_i} 1}{SPC} \qquad \text{(Equation C.19)}$$

Total Number of Backward Citations (duplicates removed)

This is the average number of backward citations for the patents in a technological domain with the duplicates removed. This measure is calculated using Equation C.20 where SPC is the simple patent count, and $BC_i$ is the number of backward citations for patent *i* and $\bigcup$ is the union of two sets and COUNT() counts the number of elements in a set. Please note that $\bigcup_{i=1}^{SPC} BC_i$ **differs from** $\sum_{i=1}^{SPC}\sum_{j=1}^{BC_i} 1$ in that the duplicates have been removed.

$$COUNT\left(\bigcup_{i=1}^{SPC} BC_i\right) \qquad \text{(Equation C.20)}$$

Total Number of U.S. Patent Classes Cited (duplicates removed)

This is the total number of US patent classes to which the backward citations belong for the patents in a technological domain. This measure is calculated using Equation C.21 where SPC is the simple patent count, $BC_i$ is the number of backward citations for patent ***i***, $UPC_{ij}$ is the **set** of US Patent classification codes for each backward citation *j* of patent *i*, and $\bigcup$ **is the union** of two sets and COUNT() counts the number of elements in a set.

$$COUNT(\bigcup_{i=1}^{SPC}(\bigcup_{j=1}^{BC_i}UPC_{ij})$$ (Equation C.21)

Total Number of International Patent Classes Cited (duplicates removed)

This is the total number of international patent classes to which the backward citations belong for the patents in a technological domain. This measure is calculated using Equation C.22 where SPC is the simple patent count, $BC_i$ is the number of backward citations for patent *i*, $IPC_{ij}$ **is the <u>set</u>** of International Patent classification codes for each backward citation *j* of patent *i*, **and** $\bigcup$ is the union of two sets and COUNT() counts the number of elements in a set.

$$COUNT(\bigcup_{i=1}^{SPC}(\bigcup_{j=1}^{BC_i}IPC_{ij})$$ (Equation C.22)

Average Number of U.S. Patent Classes Cited Per Patent

This is the average number of US patent classes to which the backward citations belong for each patent in a technological domain. This measure is calculated using Equation C.23 where SPC is the simple patent count, $BC_i$ is the number of backward citations for patent *i*, $UPC_{ij}$ is the **<u>set</u>** of US Patent classification codes for each backward citation *j* of patent *i*.

$$\frac{\sum_{i=1}^{SPC}\sum_{j=1}^{BC_i}UPC_{ij}}{SPC}$$ (Equation C.23)

Average Number of International Patent Classes Cited Per Patent

This is the average number of International patent classes to which the backward citations belong for each patent in a technological domain. This measure is calculated using Equation C.24 where SPC is the simple patent count, $BC_i$ is the number of backward citations for patent *i*, $IPC_{ij}$ is the **set** of International Patent classification codes for each backward citation *j* of patent *i*.

$$\frac{\sum_{i=1}^{SPC}\sum_{j=1}^{BC_i} IPC_{ij}}{SPC} \qquad \text{(Equation C.24)}$$

Average Number of US Patent Classes Per Patent

This is the average number of US patent classes to which each patent belongs in a technological domain. This measure is calculated using Equation C.25 where SPC is the simple patent count, $UPC_i$ is the **set** of US Patent classification codes for each patent *i*.

$$\frac{\sum_{i=1}^{SPC} UPC_i}{SPC} \qquad \text{(Equation C.25)}$$

Average Number of International Patent Classes Per Patent

This is the average number of international patent classes to which each patent belongs in a technological domain. This measure is calculated using Equation C.26 where SPC is the simple patent count, $IPC_i$ is the **set** of International Patent classification codes for each patent *i*.

$$\frac{\sum_{i=1}^{SPC} IPC_i}{SPC} \quad \text{(Equation C.26)}$$

Column 1 of Table C.1 shows the list of all of the domain patent markers that were tested in this thesis. Each of the DPMs applies to an entire set of patents that represent a TD. Column 2 shows the Pearson correlation coefficient with the TIR for each domain. Column 3 indicates whether the null hypothesis has been accepted or rejected at the 2% level. A statistically significant result should reject the null hypothesis. Column 4 shows the hypothesis that the DPM is most closely related to.

| **DPM** | **Pearson Correlation Coefficient** | **H0 (2%)** | **Hypothesis** |
|---|---|---|---|
| Simple Patent Count | 0.33 | accepted | 1: Effort |
| Average Number of Patents per Year | 0.34 | accepted | 1: Effort |
| Average Number of Forward Citations per Patent | 0.48 | rejected | 2: Breakthroughs |
| Total Number of Forward Citations (duplicates removed) | 0.44 | rejected | 2: Breakthroughs |
| Total Number of U.S. Patent Classes in Forward Citations (duplicates removed) | 0.10 | accepted | 2: Breakthroughs |
| Total Number of International Patent Classes in Forward Citations (duplicates removed) | 0.20 | accepted | 2: Breakthroughs |
| Average Number of US Patent Classes in the Forward Citations per Patent | -0.07 | accepted | 2: Breakthroughs |
| Average Number of International Patent Classes in the Forward Citations per Patent | -0.02 | accepted | 2: Breakthroughs |
| Ratio of Patents with more than 20 forward citations | 0.38 | accepted | 2: Breakthroughs |
| Total Number of Patents with More than 20 Forward Citations | 0.37 | accepted | 2: Breakthroughs |
| Average number of Forward citations within 3 years of publication per patent | 0.76 | rejected | 2: Breakthroughs |
| Average number of Forward citations within 5 years of publication per | 0.73 | rejected | 2: Breakthroughs |

| | | | |
|---|---|---|---|
| patent | | | |
| Price Ratio (3 years) | 0.39 | accepted | 2: Breakthroughs |
| NPL Ratio | 0.20 | accepted | 3: Science |
| Number of patents with >0 Citaitons | 0.39 | accepted | 3: Science |
| Ratio of patents with >0 Citaitons | 0.39 | accepted | 3: Science |
| Average Application Year | 0.53 | rejected | 4: Recency |
| Average publication year | 0.54 | rejected | 4: Recency |
| Average Processing Time Per Patent | 0.34 | accepted | 4: Recency |
| Patents in the 1970s | 0.08 | accepted | 4: Recency |
| Patents in the 1980s | 0.20 | accepted | 4: Recency |
| Patents in the 1990s | 0.32 | accepted | 4: Recency |
| Patent since the year 2000 | 0.34 | accepted | 4: Recency |
| Average Range of Backward Citation Publication Years per patent | -0.39 | accepted | 4: Recency |
| Average date of backward citation publication | 0.54 | rejected | 4: Recency |
| Average date of forward citation publication | 0.54 | rejected | 4: Recency |
| Average Range of Forward Citation Publication Years per Patent | -0.30 | accepted | 4: Recency |
| Average Age of Citations | -0.59 | rejected | 4: Recency |
| Average number of Backward Citations per patent | 0.24 | accepted | 5: Breadth of Knowledge |
| Total number of backward citations (duplicates removed) | 0.39 | accepted | 5: Breadth of Knowledge |
| Percentage of backward citations to own domain | 0.11 | accepted | 5: Breadth of Knowledge |
| TIR-Weight Backward Citations | 0.66 | rejected | 5: Breadth of Knowledge |
| Average number of US patent classes cited per patent | -0.14 | accepted | 5: Breadth of Knowledge |
| Average number of International patent classes cited per patent | -0.04 | accepted | 5: Breadth of Knowledge |
| Total number of US patent classes cited (duplicates removed) | 0.14 | accepted | 5: Breadth of Knowledge |
| Total number of international patent classes cited (duplicates removed) | 0.06 | accepted | 5: Breadth of Knowledge |
| Average number of US Patent Classes per Patent | 0.15 | accepted | 5: Breadth of Knowledge |
| Average number of International Patent Classes per Patent | -0.02 | accepted | 5: Breadth of Knowledge |
| Average Number of Inventors Per Patent | 0.27 | accepted | Other |
| Average Number of Assignees per Patent | 0.12 | accepted | Other |
| Average number of INPADOC family members per patent | 0.17 | accepted | Other |

| | | | |
|---|---|---|---|
| Average Length (in characters) of the Title per Patent | 0.59 | rejected | Other |
| Average Length (in characters) of the abstract per Patent | 0.25 | accepted | Other |
| Average Length (in characters) of the first claim per Patent | 0.03 | accepted | Other |

**Table C.1: Correlations for All 44 Domain Patent Markers and their related hypotheses**

## Appendix D: All Regression Models Tested

A number of DPMs showed a significant correlation with the TIRs and thus provided a large number of options for variables that could be included in the regression models for predicting the TIR. The models discussed in section 4.3.7.3 proved to be the most accurate, many other predictive regression models were tested. The results of those tests are shown in figure D.1.

| Model | R2 Value | Intercept | Average publication year | Average Age of Citation | Average Cited by within 3 years | Length (in characters) of Title | Average number of forward citations | % of patents with cited by over 20 | Price Index (3 years) | TIR Weighted Backward Citations |
|---|---|---|---|---|---|---|---|---|---|---|
| **A** | 0.58 | -0.26 | | | 0.16 | | | | | |
| **A p-values** | | 0.01 | | | 0 | | | | | |
| **B** | 0.64 | -31.13 | 0.02 | | 0.14 | | | | | |
| **B p-values** | | 0.05 | 0.05 | | 0 | | | | | |
| **C** | 0.72 | -34.61 | 0.02 | | 0.09 | | | | | 2.11 |
| **C p-values** | | 0.02 | 0.02 | | 0.01 | | | | | 0.02 |
| **D** | 0.72 | -34.88 | 0.02 | | 0.09 | 0 | | | | 2.12 |
| **D p-values** | | 0.02 | 0.02 | | 0.02 | 0.95 | | | | 0.02 |
| **E** | 0.59 | -0.34 | | | | 0 | | | | 0.14 |
| **E p-values** | | 0.02 | | | | 0.43 | | | | 0 |
| **F** | 0.64 | -41.63 | 0.02 | | | 0 | | | | 3.06 |
| **F p-values** | | 0.01 | 0.01 | | | 0.27 | | | | 3.72 |
| **G** | 0.64 | -0.21 | | | 0.12 | | | | | 1.9 |
| **G p-values** | | 0.02 | | | 0 | | | | | 0.04 |
| **H** | 0.66 | -0.4 | | 0.01 | 0.14 | | | | | 2.54 |
| **H p-values** | | 0.07 | | 0.32 | 0 | | | | | 0.03 |
| **I** | 0.58 | -0.19 | | 0 | 0.15 | | | | | |
| **I p-values** | | 0.37 | | 0.7 | 0 | | | | | |
| **J** | 0.6 | -0.23 | | | 0.2 | | -0.01 | | | |
| **J p-values** | | 0.01 | | | 0 | | 0.23 | | | |
| **K** | 0.6 | -0.25 | | | 0.19 | | | -0.47 | | |
| **K p-** | | 0.01 | | | 0 | | | 0.3 | | |

| | | | | |
|---|---|---|---|---|
| values | | | | |
| **L** | 0.6 | 0.39 | -0.03 | 0.44 |
| **L p-values** | | 0.12 | 0.01 | 0.52 |

**Figure D.1: Statistical Fits for Many Regressions for Predicting the TIR**

The models discussed in section 4.3.7.3 are represented by model A, B and C in figure D.1. These models show the highest $R^2$ for a given number of variables. Model D includes the length of characters in the title that was shown to be highly correlated with TIR, the inclusion of this variable does not add significantly to the predictive capability. In Model E the length of title was combined with the TIR-Weighted backward citations and additionally with average publication date in Model F, the results do not show an increase in predictive capability over models B and C. Models G through K use the strongest signal of Forward Citations with 3 years of publication as the basis of prediction and attempt to add other potentially relevant combinations of variables, most of the additional variables provide a very small improvement in predictive capability over the Cit3 variable alone. Model L incorporates the Price Index and Average age of citation and does not show an improvement over Model B.

While many of the variables did not add a significant amount of predictive capability to the models, the overall predictive power of the models tested was impressive, with all $R^2$ above 0.5 and most 0.6 or higher. The strength of these models while using a wide range of DPMs is a strong indicator in support of the related hypotheses, most notably Hypotheses 2 and 4.

# References


3D Printing Comparison Tool. (2014). *3Ders*. Retrieved from http://www.3ders.org/pricecompare/3dprinters/?tab=Details

3D Systems Viper Si2 Spec Sheet. (2014). Retrieved from http://www.3dsystems.com/products/datafiles/viper/datasheets/International/viper_si2_uk.qxd.pdf

Ahuja, G., Morris Lampert, C., & Lampert, C. M. (2001). Entrepreneurship in the large corporation: a longitudinal study of how established firms create breakthrough inventions. *Strategic Management Journal*, *22*(6-7), 521–543. doi:10.1002/smj.176

Alexander, A., & Mitchell, B. (1985). Measuring Technological Change of Heterogeneous Products. *Technological forecasting and social change*, *195*, 161–195. Retrieved from http://www.sciencedirect.com/science/article/pii/0040162585900587

Arrow, K. (1962). The Economic Implications of Learning by Doing. *The review of economic studies*, *29*(3), 155–173. Retrieved from http://www.jstor.org/stable/10.2307/2295952

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

Atkinson, K. (2008). Toward a more rational patent search paradigm. *Proceedings of the 1st ACM workshop on Patent*, 37–40. Retrieved from http://dl.acm.org/citation.cfm?id=1458582

Auerswald, P., Kau, S., Lobo, H., & Shell, K. (2000). The production recipes approach to modeling technological innovation : An application to learning by doing, *24*.

Baillie, J. (2002). *Introduction to Patent Searching.*

Balconi, M., Brusoni, S., & Orsenigo, L. (2010). In defence of the linear model: An essay. *Research Policy*, *39*(1), 1–13. doi:10.1016/j.respol.2009.09.013

Barberá-Tomás, D., Jiménez-Sáez, F., & Castelló-Molina, I. (2011). Mapping the importance of the real world: The validity of connectivity analysis of patent citations networks. *Research Policy*, *40*(3), 473–486. doi:10.1016/j.respol.2010.11.002

Basnet, S. (2014). *Textual Analysis in Patents for measuring Complexity.*

Benson, C. L., & Magee, C. L. (2012a). A Framework for Analyzing the Underlying Inventions That Drive Technical Improvements in a Specific Technological Field. *Engineering Management Research*, *1*(1), 2–14. doi:10.5539/emr.v1n1p2

Benson, C. L., & Magee, C. L. (2012b). A hybrid keyword and patent class methodology for selecting relevant sets of patents for a technological field. *Scientometrics*, *96*(1), 69–82. doi:10.1007/s11192-012-0930-3

Benson, C., & Magee, C. (2014). On improvement rates for renewable energy technologies : Solar PV , wind turbines , capacitors , and batteries. *Renewable Energy*. Retrieved from http://www.sciencedirect.com/science/article/pii/S0960148114001451

Bound, J., Cummins, C., Griliches, Z., Hall, B., & Jaffe, A. (1982). Who Does R&D and Who Patents? *NBER Working*, (June). Retrieved from http://www.nber.org/papers/w0908

Bresnahan, T., & Trajtenberg, M. (1995). General Purpose Technologies "Engines of growth"? *Journal of econometrics*, *65*, 83–108. Retrieved from http://www.sciencedirect.com/science/article/pii/030440769401598T

Brynjolfsson, E., & McAfee, A. (2011). *Race Against The Machine: How the Digital Revolution is Accelerating Innovation, Driving Productivity, and Irreversibly Transforming Employment and the Economy*. Lexington, MA: Digital Frontier Press.

Bush, V. (1945). Science: The Endless Frontier. *Transactions of the Kansas Academy of Science (1903-…*, *48*(3), 231–264. Retrieved from http://www.jstor.org/stable/10.2307/3625196

Caballero, R., & Jaffe, A. (1993). How high are Giants' Shoulders: An empirical of Knowledge spillovers and creative destruction in a model of economic growth. *NBER Macroeconomics Annual 1993, Volume 8*, *8*(January), 15–86. Retrieved from http://www.nber.org/chapters/c10998.pdf

Campbell, R. (1983). Patent trends as a technological forecasting tool. *World Patent Information*, (September 1979), 137–143. Retrieved from http://www.sciencedirect.com/science/article/pii/0172219083901345

Carlson, J. G. (1973). Cubic Learning Curves-Precision Tool for Labor Estimating. *Manufacturing Engineering and Management*, *71*(5), 22–25.

Carlson, J. G. (1976). Cubic Learning Curves-Precision Tool For Labor Estimating. *Industrial Engineering*, *8*(9), 40–47.

Christensen, C. M. (1992). Exploring the Limits of the Technology S-Curve. Part I: Component Technologies. *Production and Operations Management*, *1*(4), 334–357. doi:10.1111/j.1937-5956.1992.tb00001.x

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

Christensen, CM. (1992). Exploring the limits of the technology s-curve. Part II: Architectural Technologies. *Production and Operations Management*, *1*(4), 358–366. Retrieved from http://onlinelibrary.wiley.com/doi/10.1111/j.1937-5956.1992.tb00002.x/full

Comin, D., & Mestieri, M. (2013). Technology Diffusion: Measurement, Causes and Consequences. Retrieved from http://www.nber.org/papers/w19052.pdf

Criscuolo, P. (2006). The'home advantage'effect and patent families. A comparison of OECD triadic patents, the USPTO and the EPO. *Scientometrics*, *66*(1), 23–41. Retrieved from http://www.akademiai.com/index/p72037003qt65253.pdf

Cullinan, M. a., Panas, R. M., DiBiasio, C. M., & Culpepper, M. L. (2012). Scaling electromechanical sensors down to the nanoscale. *Sensors and Actuators A: Physical*, *187*, 162–173. doi:10.1016/j.sna.2012.08.035

D'hondt, E. (2009). Lexical issues of a syntactic approach to interactive patent retrieval. *The Proceedings of the 3rd BCSIRSG Symposium on ...*, 102–109. Retrieved from http://lands.let.kun.nl/literature/dhondt.2009.1.pdf

Dosi, G. (1982). Technological paradigms and technological trajectories. *Research Policy*, *11*(3), 147–162. doi:10.1016/0048-7333(82)90016-6

Dutton, J., & Thomas, A. (1984). Treating Progress Functions a Managerial Opportunity. *Academy of Management Review*, *9*(2), 235–247. Retrieved from http://www.jstor.org/stable/10.2307/258437

Fehrenbacher, K. (2011). Tesla CEO: I'd Bet On Capacitors Over Batteries. *GigaOm*. Retrieved November 01, 2013, from http://gigaom.com/2011/03/16/tesla-ceo-id-bet-on-capacitors-over-batteries/

Fleming, L., & Sorenson, O. (2004). Science as a map in technological search. *Strategic Management Journal*, *25*(89), 909–928. doi:10.1002/smj.384

Fontana, R., Nuvolari, A., & Verspagen, B. (2009). Mapping technological trajectories as patent citation networks. An application to data communication standards. *Economics of Innovation and New Technology*, *18*(4), 311–336. doi:10.1080/10438590801969073

Foster, R. (1986). Timing Technological Transitions. *Technology in Society*. Retrieved from http://www.sciencedirect.com/science/article/pii/0160791X85900223

Freeman, C. (1996). The greening of technology and models of innovation. *Technological Forecasting and Social Change*, *53*(1), 27–39. doi:10.1016/0040-1625(96)00060-1

Girifalco, L. A. (1991). *Dynamics of technological change*. New York: Van Nostrand Reinhold.

Goddard, C. (1982). Debunking the Learning Curve. *IEEE Transactions on Components, Hybrids, and Manufacturing Technology*, *5*(4), 328–335. doi:10.1109/TCHMT.1982.1136009

Godin, B. (2006). The Linear Model of Innovation: The Historical Construction of an Analytical Framework. *Science, Technology & Human Values*, *31*(6), 639–667. doi:10.1177/0162243906291865

Grilich, Z. (1984). *R&D, Patents and Productivity*. Chicago: University of Chicago Press.

Gruber, M., Harhoff, D., & Hoisl, K. (2013). Knowledge Recombination Across Technological Boundaries: Scientists vs. Engineers. *Manage. Sci., vol. Forthcoming*, *59*(4), 837–851. Retrieved from http://infoscience.epfl.ch/record/175202/files/MTEI-WP-2012-001-Gruber-Harhoff-Hoisl.pdf

Hall, B., & Jaffe, A. (2001). The NBER patent citation data file: Lessons, insights and methodological tools. Retrieved from http://www.nber.org/papers/w8498

Henderson, Rebecca. (1993). Underinvestment and Incompetence as Responses to Radical Innovation: Evidence from the Photolithographic Alignment Equipment Industry. *The RAND Journal of Economics*, *24*(2), 248. doi:10.2307/2555761

Henderson, Rebecca. (1995). Of life cycles real and imaginary: The unexpectedly long old age of optical lithography. *Research Policy*, *24*, 631–643. Retrieved from http://www.sciencedirect.com/science/article/pii/S004873339400790X

Henderson, RM. (1988). *The Failure of Established Firms in the face of technical change: A study of photolithographic alignment equipment*. Harvard. Retrieved from http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:The+Failure+of+Established+Firms+in+the+face+of+technical+change:+A+study+of+photolithographic+alignment+equipment#2

Jaffe, A., Fogarty, M., & Banks, B. (1998). Evidence From Patents And Patent Citations On The Impact Of NASA And Other Federal Labs On Commercial Innovation. *The Journal of Industrial …*, *XLVI*(2), 183–205. Retrieved from http://onlinelibrary.wiley.com/doi/10.1111/1467-6451.00068/full

Jaffe, A., Trajtenberg, M., & Fogarty, M. (2000). Knowledge spillovers and patent citations: Evidence from a survey of inventors. In *NBER/Sloan Project Report*. Retrieved from http://www.nber.org/sloan/sloan.pdf?origin=publicationDetail#page=27

Joho, H., Azzopardi, L., & Vanderbauwhede, W. (2010). A Survey of Patent Users. *Proceedings of the third symposium on Information interaction in context*, 13–22. Retrieved from http://ir.dcs.gla.ac.uk/~leif/papers/iiix2010-patent.pdf

Jorgenson, D., & Griliches, Z. (1967). The Explanation of Change Productivity Change. *The Review of Economic Studies*, *34*(3), 249–283. Retrieved from http://www.jstor.org/stable/10.2307/2296675

Kaplan, S. M. (1999). Discontinuous innovation and the growth paradox. *Strategy & Leadership*, *27*(2), 16–21. doi:10.1108/eb054631

Kazmerski, L. (2009). Photovoltaics at the Tipping Point: Taking Us From the Evolutionary to the Revolutionary. Retrieved from http://www.springerlink.com/index/vh4x15381 7p44786.pdf

Kodama, H. (1981). Automatic method for fabricating a three-dimensional plastic model with photo-hardening polymer. *Review of Scientific Instruments*, *52*(11), 1770. doi:10.1063/1.1136492

Koh, H., & Magee, C. L. (2006). A functional approach for studying technological progress: Application to information technology. *Technological Forecasting and Social Change*, *73*(9), 1061–1083. doi:10.1016/j.techfore.2006.06.001

Koh, H., & Magee, C. L. (2008). A functional approach for studying technological progress: Extension to energy technology. *Technological Forecasting and Social Change*, *75*(6), 735–758. doi:10.1016/j.techfore.2007.05.007

Larkey, L. (1999). A patent search and classification system. *Proceedings of the fourth ACM conference on Digital*, *1*(212). Retrieved from http://dl.acm.org/citation.cfm?id=313304

Levin, R. C., Klevorick, A. K., Nelson, R. R., & Winter, S. G. (1988). Appropriating the Returns from Industrial R&D. *Brookings Papers on Economic Activity*.

Liddle, J. A., & Gallatin, G. M. (2011). Lithography, metrology and nanomanufacturing. *Nanoscale*, *3*(7), 2679–88. doi:10.1039/c1nr10046g

Lucio-Arias, D., & Leydesdorff, L. (2009). An indicator of research front activity: Measuring intellectual organization as uncertainty reduction in document sets. *Journal of the American Society for Information Science and Technology*, *60*(12), 2488–2498. doi:10.1002/asi.21199

Lundberg, E. (1961). *Produktivitet och rdntabilitet*. Stockholm: P. A. Norstedt and S6ner.

Magdy, W., & Jones, G. J. F. (2010). PRES: a score metric for evaluating recall-oriented information retrieval applications. *and development in information retrieval*, 611–618. Retrieved from http://dl.acm.org/citation.cfm?id=1835551

Magee, C. (2012). Towards Quantification of the Role of Materials Innovation in Overall Technological Development. *Complexity*, *18*(1), 10–25. doi:10.1002/cplx

Magee, C. L. (2012). Materials Innovation in Overall Technological, *18*(1), 10–25. doi:10.1002/cplx

Magee, C. L., & Devezas, T. C. (2011). How many singularities are near and how will they disrupt human history? *Technological Forecasting and Social Change*, *78*(8), 1365–1378. doi:10.1016/j.techfore.2011.07.013

Magee, C. L., & Funk, J. L. (2014). *Exponential Change: What drives it? What does it tell us about the future?* Amazon Digital Services, Inc.

Mahdabi, P., Keikha, M., Gerani, S., & Landoni, M. (1991). *Multidisciplinary Information Retrieval. Building queries for prior-art search*. Berlin: Springer.

Margolis, R. M., & Kammen, D. M. (1999). Underinvestment: The Energy Technology and R&D Policy Challenge. *Science*, *285*(5428), 690–692. doi:10.1126/science.285.5428.690

Martino, J. (1971). Examples of Technological Trend Forecasting for Research and Development Planning. *Technological Forecasting and Social Change*. Retrieved from http://www.sciencedirect.com/science/article/pii/0040162571900035

McDonald, A., & Schrattenholzer, L. (2001). Learning rates for energy technologies. *Energy Policy*, *29*(4), 255–261. doi:10.1016/S0301-4215(00)00122-1

Mcnerney, J., Farmer, J. D., Redner, S., & Trancik, J. E. (2011). Role of design complexity in technology improvement. *Proceedings of the* …, (38). doi:10.1073/pnas.1017298108/-/DCSupplemental.www.pnas.org/cgi/doi/10.1073/pnas.1017298108

Mench, G. (1979). *Stalemate in Technology: Innovations Overcome the Depression*. Cambridge, MA: Ballinger.

Michel, J., & Bettels, B. (2001). Patent citation analysis. A closer look at the basic input data from patent search reports. *Scientometrics*, *51*(1), 185–201. Retrieved from http://www.akademiai.com/index/H74J50P35K266287.pdf

Mina, a., Ramlogan, R., Tampubolon, G., & Metcalfe, J. S. (2007). Mapping evolutionary trajectories: Applications to the growth and transformation of medical knowledge. *Research Policy*, *36*(5), 789–806. doi:10.1016/j.respol.2006.12.007

Moore, G. (1965). Cramming more components onto integrated circuits. *Electronics*, *38*(8). Retrieved from http://web.eng.fiu.edu/npala/EEE4996/Gordon_Moore_1965_Article.pdf

Moore, G. (2006). Moore's Law at 40. *Understanding Moore's law: four decades of innovation*, (Figure 2), 67–84. Retrieved from http://www.ece.ucsb.edu/~strukov/ece15bSpring2011/others/MooresLawat40.pdf

Mor-Speed Advertisement. (1928, July). *Modern Machine Shop*.

Muth, J. F. (1986). Search Theory and the Manufacturing Progress Function. *Management Science, 32*(8), 948–962. doi:10.1287/mnsc.32.8.948

Nagy, B., Farmer, J. D., Bui, Q. M., & Trancik, J. E. (2013). Statistical Basis for Predicting Technological Progress. (L. A. Nunes Amaral, Ed.)*PLoS ONE*, *8*(2), 1–7. doi:10.1371/journal.pone.0052669

Nemet, G. F. (2006). Beyond the learning curve: factors influencing cost reductions in photovoltaics. *Energy Policy*, *34*(17), 3218–3232. doi:10.1016/j.enpol.2005.06.020

Nemet, G., & Johnson, E. (2012). Do important inventions benefit from knowledge originating in other technological domains? *Research Policy*. Retrieved from https://www.lafollette.wisc.edu/publications/workingpapers/chinn2007-029.pdf

Nerkar, A. (2003). Old is gold? The value of temporal exploration in the creation of new knowledge. *Management Science*, *49*(2), 211–229. Retrieved from http://mansci.journal.informs.org/content/49/2/211.short

Nordhaus, W. (1996). Do Real-Output and Real-Wage Measures Capture Reality? The History of Lighting Suggests Not. *The economics of new goods*. Retrieved from http://www.nber.org/chapters/c6064.pdf

Nordhaus, W. (2009). The Perils of the Learning Model for Modeling Endogenous Technological Change, 1–17. Retrieved from http://www.nber.org/papers/w14638

Overview of Photolithography Stepper. (2014). Retrieved from www.cnx.org

Patsnap Patent Searching Tool. (2013). Retrieved from Patsnap.com

Peterschmitt, M. (2007). Technological progress dynamics: A preliminary application to the health care industry.

Popper, K. (1962). *Conjectures and Refutations: The Growth of Scientific Knowledge (London, 1963)* (3rd ed.). London: Routledge Classic. Retrieved from http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Conjectures+and+Refutations+The+Growth+of+Scientific+Knowledge#3

Price, D. J. (1965). Networks of Scientific Papers. *Science (New York, N.Y.)*, *149*(3683), 510–5. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/14325149

Rosenberg, N. (1982). *Inside the Black Box: Technology and Economics*. Cambridge, MA: Cambridge University Press.

Rycroft, R. W. (2006). Time and technological innovation: Implications for public policy. *Technology in Society*, *28*(3), 281–301. doi:10.1016/j.techsoc.2006.06.001

Sahal, D. (1979). *Patterns of Technological Innovation*. Beverly Hills, CA: Sage.

Sahal, Devendra. (1979). A Theory of Progress Functions. *AIIE Transactions*, (November 2013), 37–41. Retrieved from http://www.tandfonline.com/doi/full/10.1080/05695557908974396

Sahal, Devendra. (1981). *Patterns of Technological Innovation*. London: Addison-Wesley.

Schaller, R. (1997). Moore's Law: Past, present, and future. *Spectrum, IEEE*. Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=591665

Schmookler, J. (1966). *Invention and economic growth*. Cambridge, MA: Harvard University Press.

Schoenmakers, W., & Duysters, G. (2010). The technological origins of radical inventions. *Research Policy*, *39*(8), 1051–1059. doi:10.1016/j.respol.2010.05.013

Schumpeter, J. (1935). The analysis of economic change. *The Review of Economics and Statistics*, *17*(4), 2–10. Retrieved from http://www.jstor.org/stable/10.2307/1927845

Schumpeter, J. A. (1939). *Business Cycles: A Theoretical, Historical, and Statistical Analysis of the Capitalist Process*. New York: McGraw-Hill.

Schumpeter, J. A. (1942). *Socialism, Capitalism and Democracy*. London: Allen and Unwin.

Seebregts, A., Kram, T., Jan, G., & Bos, A. (2000). Endogenous learning and technology clustering : analysis with MARKAL model of the Western European energy system, *14*(1998), 289–319.

Smith, K. (1994). New directions in research and technology policy: identifying the key issues. *STEP Report R-01*.

Solow, R. (1956). A contribution to the theory of economic growth.

Solow, R. (1957). Technical Change And The Aggregate Production Function. *The review of Economics and Statistics*, *39*(3), 312–320. Retrieved from http://www.jstor.org/stable/10.2307/1926047

Sood, A., & Tellis, G. (2005). Technological evolution and radical innovation. *Journal of Marketing*, *69*(July), 152–168. Retrieved from http://www.jstor.org/stable/10.2307/30162062

Srikanth, N., & Funk, J. (2011). Geometric scaling and long-run reductions in cost: The case of wind turbines. *Technology Management Conference (* ..., (65). Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5996044

Trajtenberg, M. (1987). Patents, citations and innovations: tracing the links. Retrieved from http://www.nber.org/papers/w2457

Trajtenberg, Manuel. (1990). A Penny for Your Quotes: Patent Citations and the Value of Innovations. *The Rand Journal of Economics*, *21*(1), 172. doi:10.2307/2555502

Trajtenberg, Manuel, Henderson, R., & Jaffe, A. (1997). University Versus Corporate Patents : A Window On The Basicness Of Invention. *Economics of Innovation …*, *5*, 19–50. Retrieved from http://www.tandfonline.com/doi/full/10.1080/10438599700000006

Tushman, M., & Anderson, P. (1986). Technological Discontinuities and Organizational Environments. *Administrative science quarterly*. Retrieved from http://www.jstor.org/stable/10.2307/2392832

Twiss, B. (1979). *Management of Technological Innovation*. London: Longman.

U.S. Patent Statistics Chart Calendar Years 1963 - 2013. (n.d.). Retrieved from http://www.uspto.gov/web/offices/ac/ido/oeip/taf/us_stat.htm

USPTO. (2014). Decription of Patent Types. Retrieved from http://www.uspto.gov/web/offices/ac/ido/oeip/taf/patdesc.htm

Utterback, J M. (1974). Innovation in industry and the diffusion of technology. *Science (New York, N.Y.)*, *183*(4125), 620–6. doi:10.1126/science.183.4125.620

Utterback, James M, & Abernathy, W. J. (1975). A dynamic model of process and product innovation. *Omega*, *3*(6), 639–656. doi:10.1016/0305-0483(75)90068-7

Valentini, G. (2012). Measuring the effect of M&A on Patenting Quantity and Quality. *Strategic Management Journal*, *346*(April 2009), 336–346. doi:10.1002/smj

Wang, S.-J. (2011). The state of art patent search with an example of human vaccines. *Human Vaccines*, *7*(2), 265–268. doi:10.4161/hv.7.2.14004

Wohlers, T. (n.d.). *Wohlers Report*. Fort Collins, Colorado: Wohlers Associates.

Worstall, T. (2012). The Solyndra Report and the Real Benefits and Problems With Their Technology. *Forbes*. Retrieved from http://www.forbes.com/sites/timworstall/2012/03/30/the-solyndra-report-and-the-real-benefits-and-problems-with-their-technology/

Wright, T. (1936). Factors Affecting the Cost of Airplanes. *Journal of aeronautical sciences*. Retrieved from http://arc.aiaa.org/doi/pdf/10.2514/8.155

Yelle, L. E. (1979). The Learning Curve: Historical Review and Comprehensive Survey. *Decision Sciences*, *10*(2), 302–328. doi:10.1111/j.1540-5915.1979.tb00026.x

1. I would like to thank Khan Academic for the quick statistical refresher - if the reader also has an interest in re-learning basic regression theory visit -
2. Z-score tables are easily found with a simple google search of 'z-score table'
3. For more information on the INPADOC families, visit http://www.epo.org/searching/essentials/patent-families/definitions.html