NBER WORKING PAPER SERIES

GENERAL PURPOSE TECHNOLOGIES: "ENGINES OF GROWTH?"

Timothy F. Bresnahan

Manuel Trajtenberg

Working Paper No. 4148

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
August 1992

Prepared for the Conference on R&D and Productivity in Honour of Zvi Griliches, Jerusalem, May 1991. This paper is part of NBER's research program in Productivity. Any opinions expressed are those of the authors and not those of the National Bureau of Economic Research.

NBER Working Paper #4148
August 1992

General Purpose Technologies :
"Engines of Growth?"

ABSTRACT

Whole eras of technical progress and economic growth appear to be driven by a few key technologies, which we call General Purpose Technologies (GPT's). Thus the steam engine and the electric motor may have played such a role in the past, whereas semiconductors and computers may be doing as much in our era. GPT's are characterized by pervasiveness (they are used as inputs by many downstream sectors), inherent potential for technical improvements, and innovational complementarities', meaning that the productivity of R&D in downstream sectors increases as a consequence of innovation in the GPT. Thus, as GPT's improve they spread throughout the economy, bringing about generalized productivity gains.

Our analysis shows that the characteristics of GPT's imply a sort of increasing returns to scale phenomenon, and that this may have a large role to play in determining the rate of technical advance; on the other hand this phenomenon makes it difficult for a decentralized economy to fully exploit the growth opportunities offered by evolving GPT'S. In particular, if the relationship between the GPT and its users is limited to arms-length market transactions, there will be "too little, too late" innovation in both sectors. Likewise, difficulties in forecasting the technological developments of the other side may lower the rate of technical advance of all sectors. Lastly, we show that the analysis of GPT's has testable implications in the context of R&D and productivity equations, that can in principle be estimated.

Timothy F. Bresnahan
Stanford University/Dept. of Economics
Stanford, CA 94305-6072
and NBER

Manuel Trajtenberg
Tel-Aviv University/Dept. of Economics
Tel-Aviv 69978 ISRAEL
and NBER

1. Introduction

For over three decades now economists have known that technical change is the single most important force driving the secular process of economic growth (Abramovitz, 1956, Solow, 1957). Yet, relatively little progress has been made in unveiling the contents of the "residual" of aggregate production functions[1], or in characterizing as an economic phenomenon the notion of technical change that underlies it. Key to this gap in our understanding is the fact that technology itself (and its creation) remains by and large an empty concept in economics: as far as our analytical models go there is really no way to distinguish between, say, the advent of the microprocessor and the introduction of yet another electronic gadget.

In stark contrast to such "black box" notion of technology, economic historians emphasize the role played by some key technologies in the process of growth, such as the steam engine, electricity, and semiconductors (see Landes, 1969, Rosenberg, 1982). Anecdotal evidence aside, is there such a thing as "technological prime-movers"? Could it be that a handful of technologies had a dramatic impact on the growth potential of whole economies over extended periods of time? What is there in the very nature (technological and otherwise) of the steam engine, or the electric motor, or the silicon wafer, that make them prime "suspects" of having played such a role?

In this paper we attempt to forge a link between the details of individual technologies and the aggregate growth process. We put forward a view of innovation and growth in which there are key *technological facts* that may have far reaching consequences for the dynamic performance of the economy as a whole. The central notion is that, at any point in time, there are a handful of "generic", or "general purpose" technologies (GPT's) characterized by their pervasiveness (i.e. they can be used as inputs in a wide range of downstream sectors), and by their technological dynamism. Thus, as the GPT evolves and advances, it spreads throughout the economy, and in so doing it

---

[1]See however Denison (1962), and the series of papers in Part II and Part IV of Griliches (1988).

brings about and fosters generalized productivity gains.[2]

The presumed role of GPT's as "prime-movers" stems primarily from the workings of "innovational complementarities", meaning that the productivity of R&D in a downstream sector increases as a consequence of innovation in the GPT technology.[3,4] Thus, for example, the productivity gains associated with the introduction of electrical motors in manufacturing stem not only from reduced energy costs, but from the fact that the new energy source allowed the much more efficient (re)design of factories, taking advantage of the newfound flexibility of electric power. Thus, innovational complementarities entail the existence of a non-convexity in the underlying technology (a *vertical* externality) that magnifies and helps propagate the effect of innovation in the GPT. The sharing of the GPT among an increasing number of application sectors represent a second externality (the *horizontal* one).

Clearly, these non-convexities may speed up growth, but quite likely not up to the socially optimal rate. The reason is that they pose serious coordination problems that cannot be easily resolved in a market context. This is hardly surprising, since uncertainty and asymmetric information (which typically make coordination difficult), are in the essence of the creation of

[2]Griliches (1957) relates to hybrid corn in terms that correspond closely to our notion of general purpose technologies (in the context of agriculture): "Hybrid corn was the invention of a *method of inventing*, a *method* of breeding superior corn for specific localities." (p. 501; our emphasis). See also David (1990) for a closely related view on the role of generic technologies.

[3]We ommit here two additional forces that are thought to play a similar role: technological interrelatedness, and diffusion in conjunction with learning-by-doing. The first means that there is 'learning by inventing': the invention of a particular subtechnology in the context of a GPT lowers the costs of inventing the next one, which in turn contributes to span other subtechnologies further down the line. The second is more conventional: as the number of downstream sectors using the GPT increases the costs of producing the generalized input go down because of 'learning-by-doing', thus contributing to a self-sustained process of economy-wide growth.

[4]In defining innovational complementarities and understanding their role we were strongly influenced by Rosenberg's insightful 1979 essay, "Technological Interdependence in the American Economy," reproduced in Rosenberg (1982).

new knowledge (Arrow 1962). Moreover, time gaps and time sequences are an inherent feature of technological development, particularly in the context of GPT's (e.g. the transistor could not come before electricity, nor could interferon before DNA), and hence what would be required is coordination between agents located far from each other along the time and the technology dimensions. However, where there is potential for coordination failures there is also room for coordination, and which ultimately prevails depends upon the institutional arrangements that are developed, alongside or *in lieu* of market arrangements. Thus, looked from the vantage point of the evolution of GPT's, growth is seen to depend critically on the industrial organization details of a handful of markets, namely, those associated with the GPT.

A great deal of research has been done in recent years on the role of increasing returns in endogenous growth (going back to Romer's 1986 seminal contribution). However, many of these models regard the economy as "flat", in that they do not allow for explicit interactions between different sectors. Thus, the *locus* of technical change would not seem to matter, and hence there is no room to discuss explicitly issues of coordination, market structure and aggregate growth. Closely related, technical change is often assumed to be all-pervasive, that is, to occur with similar intensity everywhere throughout the economy. Clearly, one could not build a theory of growth that depends upon the details of bilateral market relations, when those details could in principle refer to any or all of the myriad of markets that make up the economy. By contrast, we identify here a particular sector (the GPT prevalent in each "era") that we regard as critical in fostering technical advance in a wide range of user industries, and hence presumably in "driving" the growth of the economy at large. The price that we pay though for the sharp focus is that the analysis is partial equilibrium, and hence the implications for aggregate growth are drawn by induction, rather than by explicit modelling of the whole economy.

The paper is organized as follows: the next section sets to characterize more in detail what we mean by GPT's, and brings in as illustrations the case of the steam engine and the electric motor (which deliver "continuous rotary motion"), and the case of electronic circuits, at the heart of which lies "binary logic". In order to highlight the workings of the vertical and horizontal externalities and their welfare implications, we begin by setting up in section 3 a formal model of GPT and application sectors that takes place in just "one-round", emphasizing the *two-way* appropriability problem that arises in this context (and the associated *bilateral* moral hazard). In section 4 we present the dynamic model drawing in a straightforward manner from Maskin and Tirole (1987) theory of dynamic oligopoly; we emphasize the role of technological uncertainty about the "other side" (e.g. how much user sectors know about future innovations in the GPT) in determining both the pace of technical advance of the whole cluster, and the level of the long run equilibrium. The dynamic model provides also the framework for the discussion of aggregate growth in section 5; short of a full fledge general equilibrium analysis, we do a "rent accounting" exercise, that is, we compute the rents to innovation generated as the GPT and the application sectors move step by step towards a long run equilibrium. We then relate these to a more conventional aggregate growth equation, and show how one could in principle estimate the effects of innovational complementarities, and test some of the implications of GPT's.

2. A Description of General Purpose Technologies

We think of the technologies prevalent in any given period as structured in a hierarchical pattern (i.e. as forming a sort of "technological tree"), which in the simplest case would consist just of two levels: a handful of "basic" technologies at the top (perhaps just one), and a large number of product classes or sectors that make use of the former at the bottom. Those at the top are characterized first of all by their *general purposeness*, that is, by their performing some generic function that is vital to the functioning of a large segment of existing or potential products and production systems. Such a generic function would be, for example, "continuous rotary motion," performed at first by the steam engine and later on by electrical motors; "binary logic" would the corresponding generic function for electronics, the obvious candidate GPT of our times.

The second distinctive characteristic of GPT's is their technological dynamism: continuous innovational efforts, as well as learning, increase over time the efficiency with which the generic function is performed. This may show up as reductions in the price/performance ratio of the products, systems or components in which the GPT is embodied, or as multidimensional qualitative improvements in them. As a consequence, the costs of the downstream sectors that use the GPT's as inputs are lowered, they may be able to develop better products, and moreover, further sectors will find it profitable to adopt the improved GPT, thus expanding the range of applications. Third and last, GPT's are characterized by the existence of innovational complementarities with the application sectors, in the sense that technical advances in the GPT make it more profitable for its users to innovate, and vice versa.

Of course, the process of technical advance along a given technological course will run at some point into steep diminishing returns, scientific breakthroughs will open up new technological opportunities, and hence the dominant GPT of the era will be eventually superseded. Thus, and in Schumpeter's spirit, we can think of the evolution over time of GPT's as spanning some sort of long technological waves.

2.1 Rotary motion: the steam engine and electricity

The universal character of the GPT's of the first and second industrial revolutions are relatively easy to grasp: by definition, 'work' involves the transformation of energy (be it human, animal, fossil, electrical, or nuclear), from one to the other of it various possible states, i.e. heat, motion (displacement), light, etc. It so happens that the production of great many goods involves, or could *potentially* be done, by the application of one particular type of energy transformation, namely, that which results in *continuous rotary motion*. That is precisely the generic function performed by the steam engine, and later on by the electric motor. However, it was by no means obvious that rotary motion would become a universal functionality: many manual jobs (e.g. sewing, polishing, cutting) could hardly be seen ex-ante as natural candidates for replacement by mechanical actions originating in continuous rotary motion. Moreover, in many cases the substitution did not make economic sense until the steam engine, and then the electric motor, could deliver such functionality at previously undreamt of, and continuously improving, price/ performance ratios. Once that happened, great amounts of ingenuity were expended making this functionality useful for a wide variety of industrial sectors; of course, these activities were driven by innovational complementarities.

The case of electric power provides a clear illustration of what these complementarities are all about, and give a sense of their tremendous importance in productivity growth. The first three decades of this century witnessed a steady decline in the price of electric-generated power, and constant improvements in the efficiency of electric motors. As a consequence, electric motors diffused rapidly throughout manufacturing (displacing the steam engine): whereas they accounted for less than 5% of installed horsepower in the U.S. at the turn of the century, the percentage rose to over 80% by 1930. It is widely believed that the large productivity gains registered during most of that period owe a great deal to this process of electrification. The point, however, is that the contribution of the electric motor to productivity growth goes far beyond the direct cost savings due to the spread of a cheaper power source. In the words of Rosenberg (1982)

> "The social payoff to electricity would have to include not only lower energy and capital costs but also the benefits flowing from the newfound freedom to redesign factories with a far more flexible power source. ... the steam engine required clumsy belting and shafting techniques for the transmission of power within the plant. These methods...imposed serious constraints upon the organization and flow of work, which had to be grouped, according to their power requirements, close to the energy source. ... With the advent of 'fractionalized' power made possible by electricity and the electric motor, it now became possible to provide power in very small, less costly units ... [this] flexibility ... made possible a wholesale reorganization of work arrangements and, in this way, made a wide and pervasive contribution to productivity growth throughout manufacturing. ...Machines and tools could now be put anywhere efficiency dictated, not where belts and shafts could most easily reach them." (pp. 77-78).

2.2 The 'era' of electronics

It is important to understand what 'general purposeness' means in the context of the dominant GPT of our times, namely, semiconductor technology. Once again, there is a particular functionality at the heart of this GPT, in this case binary logic. Its pervasiveness is now taken for granted, but the wide-scale application of binary logic beyond computing *per se* was by no means obvious until not long ago, and neither was the depth and breath of computing itself. Some economic activities such as accounting (broadly understood) were already conceptualized as "computing" when the integrated circuit was invented, and in fact automatic data processing in banking was one of the earliest uses of mainframe computers;[5] but who would have thought of say, the carburetion of an automobile engine as "computing"?

What accounts then for the general purposeness of electronic circuits? The workings of virtually any system and, in particular, of any electro-mechanical system, can be thought of as (and actually be broken down into) a series of steps that transform a given input into a desired outcome. Thus, a traditional watch transforms the power of the spring into an analog signal, depicting time; a washing machine transforms electrically-induced continuous mechanical traction into a series of actions involving the movement of parts, the opening and closing of valves, etc. Despite their variety, a vast majority of these intervening steps can in principle be done (or be replicated) by the application of binary logic, that is, by activating a circuit consisting of a series of binary elements (e.g. gates, flip-flops, etc.). This is a striking *technological* fact that has far reaching *economic* implications. What it says is that the enormous variety of seemingly disparate products, materials, methods of production, etc. conceal the uniformity of a few underlying technological principles; these principles, in turn, give rise to potent economic forces that would shape the (endogenous) process of technical change.

Contrary to popular perceptions, substituting binary logic for mechanical parts is in many cases extremely *inefficient*, if measured by the number of

[5]Think of the development of transactions-processing software (which was done mainly by the banking sector itself), and the associated changes in the operational procedures of the banks as complementary innovations, "enabled by" the GPT.

steps required by the former, and hence by the number of circuit components and operations involved. However, as the price and size of circuit components decrease dramatically, and as their reliability improves, it becomes eventually cost-effective to use *them* rather than the old electro-mechanical parts. And, in turn, these dramatic advances in costs, size and reliability are due to a large extent to the tremendous increases in the volume of production of *standardized* circuits, where 'learning' plays a key role. It is worth quoting extensively at this point from a classic textbook in electronics:

> "The ultimate in standardization is practical only with digital logic. High-volume mechanical parts can be made very economically, such as the $5 clock proves, but too many variations are possible...to achieve the kind of standardization we now have in digital integrated circuits. Clock gears, for example, can have any number of teeth and be any of an infinite number of sizes...Which such a large selection, it is impossible to produce standard gear components for general use in anywhere near the required volume...With digital integrated circuits standardization is easy. The logical equivalent of speed-reducing gear train in a clock is a chain of identical, standard, flip-flops, each of which reduces the speed by a factor of 2. **These flip-flops are identical to the ones used in a computer, a tape unit, or any other logic device** [our emphasis]...We thus have the key to the digital takeover of the world: standardized bargain components...The tremendous savings from using the standardized components more than offset the inefficiency of adapting the components to the application." (Blakeslee, 1975, page 4).

We may add that, even if mechanical parts could be standardized as much as integrated circuits, it is far from clear that the costs of producing them would decline as a function of volume nearly as steeply: there is something in the nature of the production process of integrated circuits that is highly conducive to continuous 'learning effects' (i.e. batch and yield), but it is hard to see what their counterpart would be in the production of e.g. gears.

Learning is just part of the story: independent scientific advances as well as massive investments in purposive R&D have contributed as much to the staggering pace of technical advance that has taken place in electronics in the last four decades. Take for example the number of individual components (e.g. transistors) that can be put on a single chip: in the early sixties an integrated circuit comprised just a handful of them, in the early eighties there were over one hundred thousand, and by the early nineties the number of components reached one million. Amazing as it sounds, the pace of advance along this dimension seems to conform indeed with 'Moore's Law', which states that the number of components that can be packed on a single chip would *double* every year (and this has been occurring for over 25 years!).

Reflecting both the purely physical aspects of these advances, and the extent of competition in semiconductors, the quality-adjusted prices of electronic devices have been declining at an unparalleled rate. Thus, for example, the price per kilobit of dynamic random access memories (one of the most common electronic components), has declined from over *four dollars* in the early seventies to less than *one cent* lately; in other words, a kilobit of memory was *5,000 times* more expensive 15 years ago than it is today. As a consequence of these quantum jumps, a personal computer today is many times more powerful and versatile than the first mainframe computers were. A simple digital watch, that can be bought today for a few dollars, was altogether a technological impossibility at the time when the Sputnik was launched. CT scanners, requiring the collection, processing and visual display of millions of pieces of information in a few seconds, are almost as much a commonplace in hospitals today as x-rays were a generation ago.

Thus the dominant technology of our times exhibits very clearly the key features of a GPT: first, it has proven to have the inherent potential for persistent and manifold technical advances along its main performance dimensions; and second, these advances impinge upon a wide range of applications which, coupled with complementary innovations by the user sectors, have brought about a reshaping of the universe of goods and services at our disposal.

3. A Single-Period Model of GPT and Application Sectors

We begin by modelling the interaction between the GPT and the AS's as a one-shot game, whereby each sector takes the technology decisions of all others as given. The main goal here is to highlight the workings of the two types of externalities, the vertical one (between the GPT and each application sector), and the horizontal one (across application sectors), and to explore their welfare consequences. We take up the dynamics in section 4, using as a framework Maskin and Tirole (1987) theory of dynamic oligopoly. The dynamic model will allow us to consider the effect of technological foresight on the long term equilibrium and, by having an explicit sequence of alternate moves, it leads quite naturally to growth.

As it turns out though the "static" case anylised in this section obtains as the limit of the dynamic model when the discount factor goes to zero. This can be interpreted in the present context as a case where each agent faces extreme uncertainty regarding the technological developments of the other players, and hence behaves as if he/she were myopic.

3.1 The Application Sectors

As suggested above, what characterizes a GPT is first of all its *generality of purpose*, that is, the fact that it performs some generic function that lies at the heart of very many actual or potential products and production systems. As an illustration figure 1 shows some of the application sectors (AS's henceforth) of the dominant GPT of our times, semiconductors: the first transistors were incorporated in hearing aids, shortly after in radios, then in television sets and computers; later on, the advent of the microprocessor brought about an explosion of new uses which has not yet abated. Many of these applications consist of entirely new products that were made possible by powerful integrated circuits (e.g. personal computers, CT scanners, camcorders), whereas others occurred in traditional products that underwent a gradual transformation as they began to incorporate integrated circuits (e.g. automobiles, civilian and military aircraft). Note that the only shared feature of the AS's of a given GPT is the fact that they purchase that GPT as an input, otherwise they may be as diverse as any sub-set of sectors in the economy.

Somewhat more formally, an application sector is defined here as one that, (*i*) is an actual or potential user of the GPT as an input; (*ii*) can earn positive returns by engaging in R&D of its own; and (*iii*) the rents it earns increase monotonically with the 'quality' of the GPT. The conditions that each AS faces in the markets for its inputs and outputs determine its short run equilibrium; we leave these in the background for now, and characterize the behavior of an AS by the objective function that it acts *as if* it maximizes,

$$\max_{T_a} \ \pi^a(w, z, T_a) - C^a(T_a) \equiv V^a(w, z), \qquad a \in A \tag{1}$$

where w is the price and z the quality of the purchased GPT, T_a the technological level of the AS itself (affecting the quality of the product it sells, and/or the efficiency of its production process), $C^a(\cdot)$ the R&D expenditures needed to reach the level T_a,[6] $\pi^a(\cdot)$ the gross payoff (or "rents") to technical advance in the a sector, and A the set of all AS's.

For example, z would include the number of transistors, the speed clock, and the word size of microprocessors used in micro-computers ("PC's"), T would comprise effective computing speed and other performance indices of PC's, and $C^a(T)$ the R&D expenditures of PC makers. Many of the examples that come to mind suggest that both z and T are likely to be vectors of large dimensions, and moreover, that many of their elements are not readily known or easily assessed (let alone anticipated) by "outsiders" to the respective sectors. While it does not matter for the formal analysis whether z and T are scalars or vectors (and hence we treat them as scalars), the informational structure may matter a great deal (see section 4).

From the above definitions it follows that $\pi^a_z > 0$, $\pi^a_{T_a} > 0$, and $\pi^a_w < 0$; as usual in this type of models we assume $C^a_{T_a} > 0$ and $C^a_{T_aT_a} > 0$. In order to focus just on the vertical links between the AS and the GPT we assume, first, that whatever the market conditions in which the AS operates, changes in π^a are perfectly correlated with changes in social surplus; and

[6]This "cost-of-inventing" function may depend on z as well, but we ignore such dependence here since it does not alter the analysis.

second, that π^a is the maximand regardless of the structure of control along the vertical chain.[7],[8] These assumptions allow us to examine the whole GPT/AS's cluster in terms of its efficiency at internalizing the externalities associated with the GPT, while ignoring the peculiarities of each end-market.

Crucially, we assume the presence of 'innovational complementarities' (henceforth just IC), formally defined as

$$\pi^a_{zT_a} = \frac{\partial^2 \pi^a(w,z,T_a)}{\partial z\ \partial T_a} \geq 0$$

with strict inequality holding when demand for the GPT input is strictly positive. In words, the value to the AS of an additional dollar of own R&D increases with the quality of the GPT input. Conversely, IC imply that a marginal improvement in z will result in higher rents the more the AS firm advances its own technology. Given w and z, the FOC for equilibrium in an AS is

$$(2) \qquad \pi^a_{T_a}(w,z,T_a) = C^a_{T_a}(T_a),$$

which implicitly defines the reaction function,

$$(3) \qquad T_a = R^a(z,\ w)$$

It is easy to show that $R^a(\cdot)$ will be upward sloping in z;[9] thus, the AS

---

[7]Those are standard assumptions in the literature on vertical integration: see for example Hart (1988) and Bolton and Whinston (1989).

[8]These conditions hold if, for example, the AS is perfectly competitive, in which case π^a stands for consumer surplus, that is, π^a is the equilibrium (gross) social payoff to technological advance.

[9]$\partial T_a/\partial z = \pi^a_{zT_a}/-(\pi^a_{T_aT_a} - C^a_{T_aT_a}) > 0$, where $\pi^a_{zT_a} > 0$ by the assumption of IC; the denominator is positive since second order conditions are assumed to hold: we have assumed already that $C^a_{T_aT_a} > 0$, and $\pi^a_{T_aT_a} < 0$ requires

has an incentive to increase its equilibrium technology level (and hence its R&D spending) in response to a quality improvement in the GPT input. This is one side of a sort of "dual inducement mechanism" mediated by IC.

Finally, we assume that for *all* $z > 0$ and $w \geq c$ the ranking of AS's according to $V^a(w, z)$ (see eq. 1) is the same, so that given z and w the "marginal" AS is uniquely determined by the smallest $V^a(\cdot)$ that fulfills $V^a(w, z) \geq 0$.[10] Call this sector n, then $n(w, z)$ is the largest *number* of sectors that find it profitable to use the GPT as an input given w and z, and $A(w, z)$ the corresponding *set* of sectors. Note that since $V^a_w(w, z) < 0$ and $V^a_z(w, z) > 0$, then

$$
(4) \qquad \begin{gathered} n_w(w, z) < 0, \qquad n_z(w, z) > 0 \\ A(w', z) \subseteq A(w, z) \quad \text{for} \quad w' \geq w, \quad \text{and} \\ A(w, z') \supseteq A(w, z) \quad \text{for} \quad z' \geq z. \end{gathered}
$$

that is, the set of using sectors expands as the quality of the GPT improves and its price goes down.

3.2 The GPT Sector

Assuming that there is no product differentiation in the GPT sector (i.e. that in every period it sets a single z, and hence a single w), gross profits in this sector (gross of R&D costs) for any w and z are,

$$
(w - c) \sum_{a \in A} X^a(w, z, T_a)
$$

where c is the (constant) marginal cost, and the demand function $X^a(\cdot)$ stemming from the AS's is (by the derivative property), $X^a(w, z, T_a) =$

only that π^a be strictly concave.

[10]In the context of a diffusion model, with the price/performance of the GPT decreasing over time, this sector will be the last to adopt.

$-\pi^a_w(w, z, T_a)$. Once again, since we want to focus on the vertical links between the GPT and the AS's, we ignore the internal structure of the GPT sector and assume that it consists of a single producer, which may or may not behave as a monopoly vis a vis the AS's. If it does then it sets price according to,

$$w^m(z, T, c) = \arg\max_w \; (w - c) \sum_{a \in A} X^a(w, z, T_a)$$

where T (without the subscript a) stands for the *vector* of T_a's. The restricted profit function is thus $\pi^g(z, T, c) \equiv (w^m - c) \sum_{a \in A} X^a(w^m, z, T_a)$, and the behavior of the sector is characterized by,

$$\max_z \; \pi^g(z, T, c) - C^g(z)$$

where $C^g(z)$ is the R&D function of the GPT (again, we assume $C^g_z > 0$ and $C^g_{zz} > 0$). The FOC is simply,

$$\pi^g_z(z, T, c) \equiv (w^m - c) \sum_{a \in A} x^a_z(w^m, z, T_a) = C^g_z(z), \tag{5}$$

which defines the reaction function,

$$z = R^g(T, c) \tag{6}$$

Notice that the optimal z is determined by the technological levels of all AS's (T is a vector), and hence upgrades in the technology of *any* of the AS's will induce an adjustment in the quality of the GPT according to,[11]

[11] We assume for simplicity that since there are many AS's, each sector does not take into account the effect of its technological upgrades on the price of the GPT, that is, even though $\partial X^a/\partial T_a > 0$, in setting the optimal T_a each AS behaves as if $\partial w(z,T,c)/\partial T_a = 0$; letting $\partial w/\partial T_a > 0$ does not change the results, only complicates the analysis.

(7) $$R^g_{T_a} \equiv \frac{\partial z}{\partial T_a} = \frac{\pi^g_{zT_a}}{-(\pi^g_{zz} - C^g_{zz})}$$

We assume that the innovational complementarities that are present in $\pi^a(\cdot)$ get transmitted to the demand function, so that $\pi^a_{zT_a} > 0 \Rightarrow X^a_{zT_a} > 0$, and therefore $\pi^g_{zT_a} = (w^m - c)\ X^a_{zT_a} > 0$.[12] Assuming that SOC hold the denominator will be positive as well, and hence (7) will be positive, that is, a technological improvement in any of the user sectors will prompt the GPT producer to engage in further R&D and upgrade the quality of the GPT. This is then the second half of the dual inducement mechanism posited above, which is mediated here by the demand function. Recalling that also $R^a_z > 0$, the innovative activities of the GPT and of the AS can be characterized as "strategic complements" (Bulow et al, 1985).

3.3 The Decentralized Equilibrium versus the Social Optimum

Assuming that the GPT and the AS's engage just in arms-length market transactions (hence ruling out technological contracting and other forms of cooperative links), we obtain the (decentralized) Nash equilibrium by solving simultaneously for (3) and (6), that is, $\{T^o, z^o\}$ is an equilibrium iff

$$T^o_a = R^a(z^o) \qquad \forall a$$

and,

$$z^o = R^g(T^o)$$

where for some AS's it may be that $T^o_a = 0$. Typically there will be multiple Nash equilibria (if the reaction functions are concave, there will be at least two, a "low" and a "high"); moreover, one can always define *constrained*

[12] This is equivalent to assuming that changes in T_a shift demand *and* marginal revenue in the same direction, which ensures that even monopoly pricing by the GPT will not prevent an upward sloping $R^g(T_a)$ function. More competitive pricing behavior by the GPT sector would require a weaker assumption.

equilibria, one for each subset $A \subseteq \hat{A}$, where $\hat{A}$ is the set of all possible AS's (participation lags or similar rigidities would make these equilibria meaningful). The plausibility of alternative equilibria is an interesting issue on its own; however, here we are interested primarily in analyzing the efficiency of different vertical arrangements vis a vis the social optimum, and hence for comparison purposes we take the "best" decentralized equilibrium, that is, the one associated with the largest A, denoted by A^0, which will be associated with the largest z^0 and T^0 by virtue of (4).

Now to the social optimum. First we impose marginal cost pricing ($w = c$), which implies $\pi^g = 0$. For any $A \subseteq \hat{A}$ the social planner's problem is,

$$(8) \qquad \underset{z, T_a}{Max} \; \{ \sum_{a \epsilon A} \pi^a(c, z, T_a) - \sum_{a \epsilon A} C^a(T_a) - C^g(z) \} \equiv S(A)$$

rendering the FOC's,

$$(9) \qquad \pi^a_{T_a}(c,z,T_a) = C^a_{T_a}(T_a), \qquad \forall a$$

$$(10) \qquad \sum\nolimits_{a \epsilon A} \pi^a_z(c, z, T_a) = C^g_z(z),$$

which implicitly define in turn the 'socially optimal reaction functions',

$$(9)' \qquad T_a = \bar{R}^a(z, c) \qquad \forall a$$

$$(10)' \qquad z = \bar{R}^g(T)$$

For a given A, the social optimum is the vector $\{z^*, T^*\}$ that fulfills $T^*_a = \bar{R}^a(z^*, c)$ $\forall a$, and $z^* = \bar{R}^g(T^*)$. Finally, the social planner chooses the optimal set A^* according to,

$$A^* = \underset{A}{arg\,max} \; S(A)$$

where S(A) is defined in (8). Note that the marginal sector here is the one with the smallest $V^a[c, \tilde{R}^g(T^*)] \geq 0$.

Proposition 1:

Assuming that $\pi^g_z(\cdot) \leq \sum_{a\epsilon A}\pi^a_z(\cdot)$, and for any pricing rule $w^m \geq w > c$, the social optimum entails higher technological levels than the decentralized equilibrium, that is, $z^* > z^o$, $T^*_a > T^o_a$ $\forall a$, and $A^o \subseteq A^*$.

(the proof is in Appendix 1). The assumption that $\pi^g_z(\cdot) \leq \sum_{a\epsilon A}\pi^a_z(\cdot)$ deserves some further elaboration: $\pi^g_z(\cdot)$ is the value of a quality increase for the AS's at the margin, whereas $\sum_{a\epsilon A}\pi^a_z(\cdot)$ is the *total* valuation of a quality upgrade. As in the case of the provision of quality by a monopoly (Spence, 1975) in all probability the two would not be the same, and that will lead to a divergence between social and private optimality. Still, the inequality need not be as assumed (it could go the other way around), but in the present context it is arguably more plausible, since it implies that the proportion of the surplus appropriated by the GPT sector does not increase with z.[13]

This granted, the reason for the divergence between the social optimum and the decentralized equilibrium lies in the complementarities between the two inventive activities, and the positive feedbacks that they generate. Thus, starting from the social optimum $\{z^*, T^*_a\}$ and reasoning "backwards", each player would want to innovate *less*: lowering z lowers each T_a (see eq. 3), which in turn means less commercial opportunity for the GPT sector, and hence lower z (we pretend that the adjustment takes place in a sequence of steps just to illustrate the point). The effects of the participation decision by applications sectors reinforces these tendencies: lower z means lower π^a's, and as some turn negative for particular sectors, the set A shrinks. This means that the market for the GPT shrinks, prompting a further cutback in z, and hence in the T_a of those applications sectors that remain active.

[13]Actually proposition 1 may hold even if the inequality is reversed, but we have not been able to characterize the range of cases for which that is so.

It is important to note that the assumption of monopoly pricing by the GPT is *not* the villain, as can be seen by considering alternative pricing mechanisms to get a better outcome. First, pick a pricing rule that gives the AS's the right incentives to innovate: the only such rule is $w = c$, which leads to no appropriability and thus no innovation in the GPT. Second, attempt to pick a pricing rule that gives the GPT the social rate of return to innovation. Clearly, no single $w(\cdot)$ would suffice, only the perfectly price-discriminating GPT monopolist would earn the social return, but that would leave zero returns to technical advance in the AS's. A fully specified technology contract could probably solve the problem (provided that it is binding, a big "if"), but that just underlines the point made here, namely, that any arms-length *market* mechanism under innovational complementarities necessarily entails private returns that fall short of social returns for either upstream or downstream innovations, under *all* plausible pricing rules.

3.4 The Vertical and Horizontal Externalities

As already suggested, the feedback machanism leading to social rates of return greater than private ones reflects two fundamental externalities. The first is vertical, linking the payoffs of the inventors of the two complementary assets, and follows from innovational complementarities. The second is horizontal, linking the interests of players in different application sectors, and is an immediate consequence of generality of purpose.

The vertical externality is closely related to the familiar problem of appropriability, except that here it runs both ways, and hence corresponds to a *bilateral* moral hazard problem (Holmstrom, 1982, Tirole, 1988). Firms in any AS and the GPT sector have linked payoffs; the upstream firm would innovate only if there is a mechanism (involving $w > c$) that allows it to appropriate some of the social returns. The trouble is that any $w > c$ implies that the private incentive for downstream innovation is too low. For appropriability in the familiar range it is clear that neither side will have sufficient incentives to innovate.

Recently, several scholars as well as industry advocates have suggested broad-based changes in government policy to increase appropriability in sectors that would qualify as GPT's (primarily semiconductors). Typically, these policy initiatives concern intellectual property protection, limits on foreign competition, and the relaxation of antitrust standards for these sectors. What our analysis suggest is that policy measures of this nature cannot be sensibly evaluated in isolation, since they would change the incentive to innovate in the GPT sector, *and* they would change the returns to complementary investments made by users of the GPT throughout the economy. What is required is a close examination of the feedbacks and tradeoffs involved, and of the comparative statics of the system as a whole.

The second externality stems from the generality of purpose of the GPT. From the vantage point of the GPT the AS's represent commercial opportunity; thus, the more AS's there are, and the larger their demands, the faster will be the rate of change in the GPT technology. From the point of view of the AS's, expanding the set A, raising T_a for any AS, or making an AS more willing to pay for the GPT makes *all* of the other AS's better off, by raising z. Yet in equilibrium each AS finds itself with too few parallel sectors, each innovating too little.[14] The point is that, from their perspective, z is a public good while R^g is the (common) fixed cost needed to produce that good; however, attempts to cover such costs with transfer prices impose a tax that discourages innovation.

The horizontal externality can illuminate some issues in the economics of technology connected with the role of large, predictable demanders, which are in turn related to policy. It is often claimed that the procurement policy of the U.S. Defense Department "built" the microelectronics-based portion of the electronics industry in the US during the fifties and sixties. Obviously, the presence of a large demander changes the conditions of supply, and this may benefit other demanders. However, the important point here is that such a

[14]Note that this issue arises above and beyond the multiple equilibrium problem, since we have assumed that the "best" Nash equilibrium is the one that holds; in particular, the economy is not trapped at z=0, T=0 by a failure to realize mutually profitable opportunities.

demander had a high willingness to pay for components embodying z well outside current technical capabilities, and was willing to shoulder part of the risk, primarily by procurement assurances; in so doing it may have indeed set in motion (and sustained for a while) the virtuous cycle mediated by the horizontal externality.

However, it is only a coincidence that the horizontal spillouts came from the demand activity of a government entity: in the same technology, large private demanders such as the Bell System and IBM, contributed directly to the development of fundamental advances in microelectronics. Earlier GPT's displayed similar patterns, as for example in Rosenberg's (1982) description of the importance of improvements in the quality of materials for 19th century U.S. growth: much of the private return to improvements in material sciences (and engineering) came from a few key sectors, notably transportation. The need to build steel rails for the railroad, and to contain steam in both railroads and steamships, provided a kind of demand parallel to that of the government body noted above. Focused on improvements in inputs that press the technical envelope, having high willingness to pay because they themselves are making changes which are large relative to the size of the economy, such demanders provide substantial horizontal spillouts to the extent that the technical progress they induce is generally useful.

These examples seem to suggest that the "triggers" often take the form of exogenous forces that shift the rate of return to GPT technology. Thus in the 19th century the importance of certain sectors (e.g. transportation) as driven by the economic development of the country may have been the key. In the post WW2 era the onset of the cold war, and the "social contract" implied by the government procurement policy that followed may have played a similar role. In each case, the positive feedback aspects of GPT and related AS developments then took over, generating very large external effects, and unleashing a process that played out for decades.

3.5 Externalities and Technological Contracting

Clearly, the vertical and horizontal externalities offer a strong motive for breaking away from the limitations of arms-length market transactions, by increasing the degree of cooperation and explicit contracting between AS's and the GPT, and between the AS's themselves. To illustrate, consider the case whereby any two agents can form an arbitrary, binding technology contract, be it the GPT sector and an AS, or a pair of AS's. It is easy to see that in the former case they will pick z and T_a to maximize $(\pi^a + \pi^g)$; in the latter, they will pick the two T_a's to maximize the sum of the two applications sectors' payoffs. The result of either such contract will be that z and T_a will be larger for *all* applications sectors: the set A can expand as a consequence of the contract, but not shrink. Payoffs will be larger for the GPT sector, and for all AS's not party to the contract as well. Note however that the activity of forming binding technology contracts is subject to the same externality as the provision of technology itself. Just as every AS would like to see other AS's advancing their own technology, so too would each sector like to see others making technology-development contracts with the GPT. Clearly, lack of enforceability, as well as imperfect technology forecasting may seriously limit the practical importance of contracting.

Recent events in the computer and telecommunications markets show how pervasive yet complex the motive for technological cooperation can be. For a long period, each market was characterized by the presence of a dominant firm (IBM and AT&T), which could take a leading role in the determination not only of its own technology, but in the encouragement of complementary developments in or for applications. The changing conditions of competition in both markets have removed the obvious enforcer of implicit technology contracts. Now, technical progress in the GPT part of both computing and telecommunications is diffused across quite a few firms, and the mechanisms for technology contracting have changed accordingly. "Strategic alliances," participation in formal standards-setting processes, consortia, software "missionaries," and the systematic manipulation of the trade press, have all emerged as standard management tools in microelectronics-based industries. These mechanisms permit both revelation of the likely direction of technical advance within particular technologies, and encouragement of complementary innovations. Yet they probably fall short of offering the means to internalize the bulk of the externalities discussed above.

4. The Dynamics of General Purpose Technologies

In previous sections we assumed that the whole process takes place in just "one round", and that allowed us to discuss the two main externalities associated with GPT's in a relatively simple fashion. However, in order to examine the implications of GPT's for growth, we need to formulate explicitly a dynamic process by which the innovational efforts of the GPT and the AS's unfold and interact over time. A suitable framework for that purpose is the theory of dynamic oligopoly as developed by Maskin and Tirole (1987) (henceforth M&T), which centers around the concept of Markov Perfect Equilibrium (MPE). In what follows we sketch the model and (re)state the main results from M&T in terms of GPT's and AS's.

Denote by $\pi^a(z_t, T_t)$ the instantaneous profit function of the AS, and by $\pi^g(z_t, T_t)$ that of the GPT (for simplicity we assume that w is fixed). The GPT and the AS are assumed to move in alternate periods of fixed lenght τ, which in the present context has a natural interpretation, namely, it is the lenght of time it takes to develop the "next generation" (either of the GPT or of the AS), given that the other side has already developed its current technology. Thus, the quality level of the GPT at time $t-1$ is z_{t-1}, and it remains constant for the next two periods (i.e. for a lenght of time of 2τ). Given z_{t-1} it takes the AS τ to develop its technology up to level T_t, and once that has materialized it takes again τ for the GPT to develop its next generation, which will be marketed in period $t+1$ and will exhibit quality z_{t+1}. We refer for simplicity to *the* AS but actually we mean *all* AS's, which are assumed to move simultaneously every other period. Thus we abstract from the process of diffusion of the GPT among AS's, but otherwise the analysis is the same whether the side interacting with the GPT is made up of one or many sectors.

An obvious difference between this and the original M&T's formulation is that in here the firms involved are not oligopolists competing in the same market; they are instead vertically related, with their innovational efforts being "strategic complements". However, from a formal point of view that is just a technicality: what counts is that the cross derivatives of the payoff functions of all players do not vanish. Notice also that whereas in the original oligopoly context the assumption of a sequence of alternate moves (and consequent two-period commitments) constitutes an awkard feature that is not easy to justify, in the present context it is a natural modelling scheme that stems at least to some extent from technological imperatives.

At time t each firm maximizes,[15]

$$(11) \qquad \sum_{s=0}^{\infty} \delta^{s} \pi^{i}(z_{t+s}, T_{t+s}), \qquad i= a, g$$

where $\delta = \exp(-r\tau)$ is the discount factor, and r the interest rate. Define a dynamic reaction function for Markov strategies (i.e. dependant only on the payoff-relevant state and not on history) for the AS as $T_t = R^a(z_{t-1})$, and similarly for the GPT, $z_t = R^a(T_{t-1})$. The pair (R^a, R^g) form a MPE iff there exist valuation functions (V^i, W^i), i=a,g, such that (for the AS),

$$V^a(z) = \max_T [\pi^a(z, T) + \delta W^a(T)]$$

$$R^a(z) \text{ maximizes } [\pi^a(z, T) + \delta W^a(T)]$$

$$W^a(T) = \pi^a[R^g(T), T] + \delta V^a[R^g(T)]$$

and similarly for the GPT. It is easy to show that the reaction functions will be in this case upward sloping, since the cross-derivates of the π^i's, π^i_{zT}, are positive (because of innovational complementarities).

[15]We abstract for the time being from the cost side, since its inclusion does not affect the substance of the analysis. Development costs correspond here to what M&T refer to as "adjustment costs" (of changing outputs in their context) - see section 5.

M&T prove that, for any discount factor δ, (a) there exists a unique linear MPE which is dynamically stable; and (b) the equilibrium (steady state) values of the decision variables (z^e, T^e) equal the static Cournot-Nash equilibrium when $\delta = 0$, and grow with δ. An equivalent way of phrasing (b) is that the (dynamic) reaction functions coincide with their static (Cournot) counterparts as δ goes to zero.[16]

This proposition has highly revealing implications in our context. The discount factor δ can be interpreted here as a measure of the difficulty in forecasting the technological developments of the other side: the smaller δ is, the more difficult it is for the AS to anticipate the future qualities of the GPT, and viceversa.[17] Technological forecasting, in turn, depends upon a variety of institutional arrangements that may facilitate or hinder the flow of credible technological information between the GPT and the AS's. Thus, part (b) of the proposition implies that such arrangements may have far reaching consequences for the actual innovational efforts of the sectors involved: the more "cooperative" they are in terms of informational exchanges, the higher the ultimate equilibrium levels (z^e, T^e) will be, and moreover, the larger the values $\{z_t, T_t\}$ will be at each step in the sequence leading towards the steady state (see Figure 2). Larger values at each step will translate in turn into faster aggregate growth, provided that in the process the GPT diffuses throughtout a large number of sectors (see next section).

In the limit ($\delta = 0$), it is altogether impossible for any player to forecast the next technological developments of the others, and hence it will have no choice but to behave as if it were myopic, that is, to decide on each move assuming as it were that the others will stay put. In other words, sophisticated forward looking planning coupled with extreme uncertainty is indistinguishable behaviorally from shortsightedness with or without complete

[16]M&T prove the proposition for the special case of quadratic profit functions; Dana and Montrucchio (1986) generalized the proof for any concave payoff function; see also Dana and Montrucchio (1987).

[17]This is of course a shortcut to the explicit modelling of technological uncertainty, which would involve games of incomplete information.

information. Thus, the (static) Cournot-like reaction functions can be seen in this context as generating an actual sequence of moves, that cannot be dispensed of with the traditional argument of inconsistensy. This is a very useful feature, since in the present context it is easy to derive Cournot reaction functions for virtually any payoff functions.

As suggested above the point is that δ is not to be taken as a given, but rather it is a function of the industrial organization features of the market for the GPT. One way to think of it is as follows: suppose that τ is the required overall development time of each "new generation" of both the GPT and the AS;[18] however, assume now that a proportion $(1 - \theta)$ of the development can be done before the other side has completed its development (which implies of course that a proportion θ *has to be done* afterwards). Thus, for example, firms developing new personal computers know that the next generation of Intel's microprocessors is going to be the 586, that it is due in late 1992, that it is expected to have 2 million transistors and at least twice the 486's performance (see table 1). On that basis they may be able to do part of the R&D for the next generation of personal computers that will incorporate the 586, but not all: some of the development process requires that they actually get hands on the 586, examine it, test it in various configurations, etc. How much they can develop prior to the actual appearance of the 586 depends *inter alia* upon the degree of detail of the technological information that they manage to obtain, the extent to which Intel is willing to make them privy of the development process, etc.

The reverse conditioning is perhaps less obvious but not less important: to continue with the same example, Intel has been developing parts and circuits for personal computers (other than microprocessors) even though "neither line is profitable as chips, but through them Intel gains insight into trends: Knowing what needs to go on a board this year helps it determine what should go into microprocessors next year" (Business Week, April 29, 1991,

[18]The period length τ can also be endogenized, i.e. it can be made a function of R&D, a strategic variable; one can easily allow also for differences in τ across sectors.

page 55). This is true to various degrees as one goes down the "technological tree": thus, software developers need to actually have the new operating systems in order to develop software for them; in order to write new operating systems one needs to get hands on the (new) personal computers that will be use them, and so forth.

Thus, the "effective" lenght of a period for our purposes (i.e. for eq. 11) is $\tau^* \equiv \theta\tau$. We think of θ as having an upper and a lower bound: if the relationship between the GPT and the AS takes the form of arms-lenght market transactions, with no coordination of any kind between them (i.e. no intended exchange of technological information), then $\theta = \bar{\theta}$, which can as well be normalized to 1. On the other hand, if the industrial organization features of these sectors are such that all technologically relevant information flows freely between the two players, then $\theta = \underline{\theta}$ (we conjecture that $\underline{\theta} > 0$, but that's a detail). There is therefore a range $\underline{\theta} \leq \theta \leq \bar{\theta}$ that maps a corresponding institutional/organizational spectrum; moreover, collective action presumably can change the prevailing θ, thus affecting the present and future pace of innovation.

Clearly, the scope for coordination in the above sense increases with the number and range of AS's (and so does the loss in a case of failure to coordinate). Thus for example an improvement in the ability of the microcomputer industry to forecast technological advances in microprocessors may speed up the use of microlectronics in *cars*, hence foster larger improvements in cars themselves, stimulate the demand for chips and hence encourage their further development, and so forth.

5. Superadditivity and Growth

So far the analysis has been *partial* equilibrium, and we intend to keep it that way. Nevertheless, one can examine the impact of GPT's on aggregate growth (albeit in a limited fashion), by looking at the rents generated along the process leading to a long-run Markov Perfect Equilibrium. The idea is simply that growth can be thought of as a process of rent creation (that is, as a process generating ever increasing *total* returns to factors), and hence one can learn about the rate of growth of alternative regimes by examining the flow of rents under each regime. That is, we will do partial equilibrium *rent accounting*, rather than the more conventional general equilibrium *growth accounting*.

Recall that it is the generality of purpose of the GPT that allows us to talk about aggregate growth in this context: the leading GPT of each "era" eventually diffuses through a very large number of AS's and spans new ones, so that the [GPT/AS's] cluster ends up accounting for a large portion of the economy as a whole. Thus, the aggregate rate of growth, and the extent of the concomitant rent-creation will depend on the rate of advance in z and on the complementary innovational efforts by the AS's. As suggested above, these in turn may well hinge on how well the GPT sector and the AS's manage to "coordinate" their innovational plans, in the sense of devising mechanisms to facilitate the flow of technological information and forecasts between them.

Let us examine now the magnitude of the rents generated as the process unfolds. Suppose that an external shock affects favorably the GPT sector (e.g. an exogenous innovation that lowers the costs of upgrading z), disturbing the present equilibrium and triggering a sequential adjustment process that will lead to a new and higher MPE.[19] Consider the first two steps of the sequence: at first the GPT producer increases its quality by Δz, and then each AS upgrades its technology by $\Delta T = R^a(z+\Delta z) - R^a(z)$. The (gross) incremental rents to the GPT consist of the sum of three parts (we do it for the moment just for one AS):

---

[19] Suppose that the dynamic reaction function of each sector is also a function of a random variable ω_i, with $\partial R^i/\partial \omega_i > 0$, $i=g,a\epsilon A$. For $\omega_i = \omega_i^0$ the (starting) equilibrium is $\{z_o, T_o\}$; at $t=0$ an external shock occurs to say, the GPT, such that $\omega_g' > \omega_g^o$. The first step in the sequence leading to a new equilibrium will be $z_1 = R^g(T_0, \omega_g')$, followed by $T_2 = R^a(z_1, \omega_a^o)$, and so forth.

(*i*) Direct own effect:[20]

$$\int_{z}^{z+\Delta z} \pi_z^g(v,\ T)\ dv \ \equiv\ \pi_z^g\ \Delta z$$

(*ii*) Feedback effect ("demand spillout"):

$$\int_{T}^{T+\Delta T} \pi_T^g(z,\ u)\ du \ \equiv\ \pi_T^g\ \Delta T$$

(*iii*) Joint effects arising from IC's ("super-additivity"):

$$\int_{T}^{T+\Delta T}\int_{z}^{z+\Delta z} \pi_{zT}^g(v,\ u)\ dv\ du \ \equiv\ \pi_{zT}^g\ \Delta z\ \Delta T$$

The incremental rents to the AS can be decomposed in a similar fashion, that is, (*i*)' direct own effect: $\pi_T^a\ \Delta T$, (*ii*)' feedback effect ("pecuniary externality"): $\pi_z^a\ \Delta z$, and (*iii*)' joint effects arising from IC's ("super-additivity"): $\pi_{zT}^a\ \Delta z\ \Delta T$. Substracting the R&D expenditures of each sector (C^g for the GPT and C^a for the AS), we can write the *total net* increments in rents as,

$$\text{(12)}\quad \Delta\Pi = \underbrace{[\pi_z^g\ \Delta z + \pi_T^a\ \Delta T]}_{\text{direct/own}} + \underbrace{\pi_T^g\ \Delta T}_{\text{demand spillout}} + \underbrace{\pi_z^a\ \Delta z}_{\text{pecuniary externality}} + \underbrace{(\pi_{zT}^g + \pi_{zT}^a)\ \Delta z\ \Delta T}_{\text{super-additivity}} - \underbrace{C^g - C^a}_{\text{R\&D costs}}$$

In order to express these rents in terms that translate more readily into empirical categories, we assume that the "innovation function" for the GPT

[20]The derivatives in the expressions of the form $\pi_j^i\ \Delta j$ (i = g,a, j = z,T) should be understood as *average* derivatives over the relevant range (e.g. $z+\Delta z$), and likewise for $\pi_{zT}^i\ \Delta z \Delta T$.

sector (i.e. the inverse of the R&D-cost function $C^g(z)$) takes the form,

(13) $$z_t = \varphi(C_t^g) + z_{t-1}$$

where C_t^g are the R&D expenditures in period t (we associate each "step" in the sequence with a time period), and similarly for the AS sectors, $T_t = \phi(C_t^a) + T_{t-1}$. We can then write[21]

$$\pi_z^g \,\Delta z = (\partial\pi^g/\partial C^g)\cdot C^g$$

and the same for the AS, $\pi_{T_a}^a \cdot \Delta T_a = (\partial\pi^a/\partial C^a)\cdot C^a$. Thus, the components of (12) can be written as follows (this time for *all* AS's),

(*i*) direct own returns: $[(\partial\pi^g/\partial C^g) - 1]\; C^g + \sum_{a\epsilon A}[(\partial\pi^a/\partial C^a) - 1]\; C^a$

(*ii*) spillovers: $\sum_{a\epsilon A}(\partial\pi^g/\partial C^a)\; C^a + \sum_{a\epsilon A}(\partial\pi^a/\partial C^g)\; C^g$

(*iii*) super-additivity: $\sum_{a\epsilon A}[(\partial^2\pi^a/\, \partial C^a\, \partial C^g) + (\partial^2\pi^g/\, \partial C^a\, \partial C^g)]\; C^a\, C^g$

In a competitive economy we would expect (*i*) to be about the same as the return to other investments; on the other hand, the rents stemming from the

[21]From (13) it follows that $\Delta z_t = \varphi(C_t^g)$, and $\Delta T_t = \phi(C_t^a)$; for example, $\Delta z_t = (C_t^g)^{1/2}$ (or, more generally, $[C_t^g]^a$, $a<1$), in which case we get the adjustment costs model as in Masking and Tirole (1987). In equilibrium $C^i = 0$ (which is the baseline), and hence $\Delta C_t^i = C_t^i$. Thus, $\Delta z_t = (\partial z/\partial C_t^g)\cdot C_t^g$, and $\Delta T_t = (\partial T/\partial C_t^a)\cdot C_t^a$. Writing the payoff functions as $\pi^i[z_t(C_t^g,\; z_{t-1}),\; T_t(C_t^a,\; T_{t-1})]$, i=a,g, we get: $(\partial\pi^g/\partial z)\cdot\Delta z_t = \partial\pi^g/\partial C_t^g)\cdot C_t^g$, and similarly for the AS, $\partial\pi^a/\partial T)\cdot\Delta T_t = (\partial\pi^a/\partial C_t^a)\cdot C_t^a$.

externalities (*ii*) and (*iii*) may be of any size, depending both on purely technological factors (e.g the strenght of IC), and on institutional and behavioral factors, which determine the steepness of the dynamic reaction functions, and hence the magnitude of each "step" along the adjustment path. To simplify the notation, define

$$\beta_g: \ (\partial\pi^g/\partial C^g) + \textstyle\sum_{a\epsilon A}(\partial\pi^a/\partial C^g) - 1$$

$$\beta_a: \ (\partial\pi^a/\partial C^a) + (\partial\pi^g/\partial C^a) - 1$$

$$\beta_{ga}: \ (\partial^2\pi^a/ \partial C^g \ \partial C^a) + (\partial^2\pi^g / \partial C^g \ \partial C^a)$$

Collecting terms, equation 12 becomes,

$$\Delta\Pi = \beta_g C^g + \textstyle\sum_{a\epsilon A}\beta_a C^a + C^g\cdot\textstyle\sum_{a\epsilon A}\beta_{ga} C^a$$

Suppose that the AS's can be aggregated, in the sense that $\beta_a = \beta_A + \bar{\beta}_a$, $\Sigma C^a\bar{\beta}_a \cong 0$, and similary for the β_{ga}'s. Then,

$$\Delta\Pi = \beta_g C^g + \beta_A\cdot\textstyle\sum_{a\epsilon A} C^a + \beta_{gA} (C^g\cdot\textstyle\sum_{a\epsilon A} C^a) \tag{13}$$

In order to link these rent increments to a conventional growth framework, consider Griliches' (1973) formula for total factor productivity growth:

$$f = \lambda + \alpha k,$$

where λ is the rate of autonomous technical change, k is the rate of growth of the stock of knowledge K, and α is the marginal product of K. This can also be written as $f = \lambda + \rho RD/Q$, where ρ is the rate of return to R&D, RD is net aggregate investment in R&D, and Q total output.

Clearly, $\Delta\Pi$ as written in (13) is equivalent to ρRD, and therefore the rate of productivity growth at time t can be expressed as $f_t = f(\lambda, \Delta\Pi_t)$. Putting it in terms of a linear regression,[22]

$$(14) \qquad f_t = \lambda + \beta_g C_t^g + \beta_A \sum_{a\epsilon A} C_t^a + \beta_{gA} (C_t^g \cdot \sum_{a\epsilon A} C_t^a) + \epsilon_t$$

which is in priciple an estimable equation, that may allow us to test some of the empirical implications of GPT's. First, note that β_g comprises the net return to own R&D that accrues to the GPT producer, plus the *sum* of the pecuniary externalities bestowed on all AS's. On the other hand, β_A includes the return to own R&D accruing to the *average AS*, plus the *average spillout* from AS's to the GPT (due to the fact that upgrading the technology of any AS increases the demand for the GPT). Thus, we expect that $\beta_g >> \beta_A$, that is, a dollar of R&D spent in the GPT sector would bring in equilibrium higher total returns than a dollar spent in any particular application sector. If so, aggregate productivity growth would no longer depend upon *aggregate* R&D (as in the traditional framework) but upon the *distribution* of R&D between the GPT and the AS's.[23]

Second, notice that β_{gA} is a measure of the strenght of innovational complementarities, and hence the force driving the *endogenous* growth process. Thus if GPT's work as posited here we would expect that $\beta_{gA} > 0$. Moreover, in the course of a "GPT era" we would expect β_{gA} to be higher at first, and to decline in later stages; in fact, the shrinking of β_{gA} is what should herald the end of the role of the GPT as an "engine of growth" in its era.

---

[22]For a related empirical study of R&D spillovers see Jaffe (1986); see also Griliches (1991) for a comprehensive survey of related work.

[23]It may also depend upon the distribution of R&D across the AS's themselves, if these cannot be aggregated as done above.

From another angle, if (14) could be estimated for a cross section of countries one could in principle test the hypothesis that more cooperation and/or better mechanisms for the transmision of technological information implies both larger β's and higher levels of R&D spending, and therefore a positive correlation between the two.

6. Concluding Remarks

Our analysis shows that the characteristics of GPT's imply a sort of increasing returns to scale phenomenon, and that this may have an important role to play in determining the rate of technical advance, and hence the rate of growth of the whole economy. On the other hand this phenomenon makes it difficult for a decentralized economy to fully exploit the growth opportunities offered by evolving GPT's. In particular, if the relationship between the GPT and its users is limited to arms-lenght market transactions, there will be "too little, too late" innovation in both the GPT and the application sectors. Likewise, difficulties in forecasting the technological developments of the other side may lower the rate of technical advance of all sectors. Lastly, we show that the analysis of GPT's has testable implications in the context of R&D and productivity equations, that can in principle be estimated.

In future work we intend to follow several tracks: first, we would like to do econometric work at the aggregate level as outlined in section 5; aside from some challenging data problems (e.g. how to identify all or most of the AS's, and obtain data on their R&D?) this would require a much tighter formulation of the equations to be estimated, clarifying at the same time how they relate to more conventional studies of R&D spillovers. Second, we would like to do micro-level studies, aimed at estimating "technological value added": how much of the gains from innovation registered in markets for final products (i.e. the markets for the AS's) are "due to" technological advances in the AS's themselves, as opposed to stemming from innovations in the GPT incorporated in the AS's; in our notation the issue is estimating and comparing π^a_Z versus π^a_T. We have collected extensive data on microcomputers, which may allow us to carry out this type of study.

Third, historical studies of GPT's and "institutions" (in the broad sense): the intention would be to examine the historical evolution of particular GPT's and of the institutions coupled with them, using our conceptual framework in trying to understand their joint dynamics. In particular, we would like to assess the extent to which specific institutions facilitated or hindered the GPT's in playing out their presumed roles as "engines of growth". A key hypothesis is that institutions display much more inertia than leading technologies, and hence as a GPT era comes to a close and new GPT's emerge, an economy may "get stuck" with the wrong institutions, that is, those that enable the previous GPT to advance and carry the AS's, but that may prove inadequate to do as much for the new GPT.

To sum up, the main goal of this paper has been to suggest a way of thinking about technical change, that focuses on the interface between the characteristics of key technologies and the features of the markets for them. It is thus an attempt to look carefully inside the "black box" of technology, inspired by history and aided by formal modelling, while seeking to unveil the links between the stylized facts of technology and the institutions surrounding it. Since at any point in time there are countless "technologies", this approach is useful only in so far as it can identify at the outset a small subset of technologies that are of particular economic relevance, and characterize them tightly. The notion of general purpose technologies put forward here fulfills that role, but that is certainly just one possible abstraction in this vein, there may be other interesting and useful characterizations as well.

References

Abramovitz, M. "Resource and Output Trends in the United States since 1870". American Economic Review Papers and Proceedings, May 1956.

Arrow, K.J. "Economic Welfare and the Allocation of Resources for Inventions," in R. Nelson (ed.) The Rate and Direction of Inventive Activity, Princeton University Press, 1962.

Blakeslee, Thomas R., Digital Design with Standard MSI and LSI. New York: John Wiley & Sons, 1975.

Bolton, P. and Whiston, M.D. "Incomplete Contracts, Vertical Integration, and Supply Assurance". Harvard University, mimeo, December 1989.

Bulow, J., Geanakoplos, J. and P. Klemperer "Multimarket Oligopoly: Strategic Substitutes and Complements". Journal of Political Economy, 1985 (93), pp. 488-511

Dana R.A. and Montrucchio, L. "Dynamic Complexity in Duopoly Games". Journal of Economic Theory, 1986 (40), pp. 40-56.

Dana R.A. and Montrucchio, L. "On Rational Dynamic Strategies in Infinite Horizon Models Where Agents Discount the Future". Journal of Economic Behavior and Organization, 1987 (8), pp. 497-511.

David, P.A. "The Dynamo and the Computer: An Historical Perspective on the Modern Productvity Paradox". American Economic Review Papers and Procedings, 1990, pp. 355-361.

Denison, E. "United States Economic Growth". Journal of Business, August 1962.

Griliches, Z. "Hybrid Corn: An Exploration in the Economics of Technological Change". Econometrica, 1957(25), pp. 501-522.

Griliches, Z. "Research Expenditures and Growth Accounting". In B.R. Williams, ed. Science and Technology in Economic Growth. London: Macmillan, 1973, pp. 59-95.

Griliches, Z. Technology, Education, and Productivity. New York: Basil Blackwell, 1988.

Griliches, Z. "The Search for R&D Spillovers". National Bureau of Economic Research, WP 3768, 1991

Hart, "Incomplete Contracts and the Theory of the Firm". Journal of Law, Economics and Organization, Spring 1988.

Holmstrom, B. "Moral Hazard in Teams". Bell Journal of Economics, 1982 (13), pp. 324-340.

Jaffe, A.: "Technological Opportunity and Spillovers of R&D: Evidence from Firms' Patents, Profits and Market Value," American Economic Review, 1986 (76), pp. 984-1001.

Landes, D. The Unbound Prometheus. Cambridge: Cambridge University Press, 1969.

Maskin, E. and Tirole, J. "A Theory of Dynamic Oligopoly, III: Cournot Competition." European Economic Review, 1987(31), pp. 947-968.

Romer, P. "Increasing Returns and Long-Run Growth," Journal of Political Economy, 1986.

Rosenberg, N. Inside the Black Box: Technology and Economics. Cambridge: Cambridge University Press, 1982.

Spence, M. "Monopoly, Quality, and Regulation". Bell Journal of Economics, 1976(6), pp. 417-429.

Solow, R. "Technical Change and the Aggregate Production Function". Review of Economic and Statistics, 1957.

Tirole, J. The Theory of Industrial Organization. Cambridge: MIT Press, 1988.

Appendix 1: Proof of Proposition 1

Compare eqs. (2) and (9), which correspond to an AS's private and social equilibrium conditions respectively:

$$(2) \qquad \pi^a_{T_a}(w,\ z,\ T_a) = C^a_{T_a}(T_a)$$

$$(9) \qquad \pi^a_{T_a}(c,\ z,\ T_a) = C^a_{T_a}(T_a)$$

The two equations are identical, except for the fact that $w > c$, and hence the LHS of (2) is smaller than the LHS of (9), since by assumption $X^a_{T_a} = -\pi^a_{wT_a} > 0$. Thus, $R^a(z,\ w) < R^a(z,\ c) = \tilde{R}^a(z,\ c)$ for all z, and in particular,

$$(A1) \qquad T_a' \equiv R^a(z^*, w) < \tilde{R}^a(z^*,\ c) = T_a^*$$

Consider now the FOC for private (eq. 5) and social (eq. 10) optimum in setting z,

$$(5) \qquad \pi^g_z(z,\ T,\ c) \equiv (w^m - c) \sum_{a\epsilon A} X^a_z(w^m, z, T_a) = C^g_z(z)$$

$$(10) \qquad \sum_{a\epsilon A} \pi^a_z(c,\ z,\ T_a) = C^g_z(z),$$

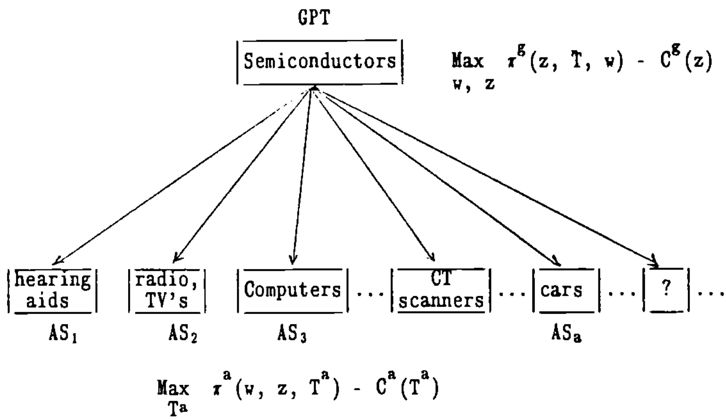
The RHS of both is the same, but the LHS of (6) is smaller than that of (10) by assumption. Thus, $R^g(T) < \tilde{R}^g(T)$ for all T and, in particular

$$(A2) \qquad z' \equiv R^g(T^*) < \tilde{R}^g(T^*) = z^*$$

Clearly, z^* cannot be part of a decentralized equilibrium (DE) since that requires it to be a fixed point, whereas by (A1) and (A2) $z^* < R^g[R^a(z^*)]$, and similarly for T^*. Relying on the same argument one can show that $\{z', T'\}$ cannot be a DE either. Assuming that SOC hold for the DE and hence that

$R^g_{T_a} < 1/R^a_z < 1$, then is clear that if $\{z^o, T^o\}$ is a DE then $z^o = R^g[R^a(z^o)] < z' < z^*$, and likewise $T^o_a = R^a[R^g(T^o_a)] < T'_a < T^*$. As to the equilibrium number of sectors, it is clear from (4) that $n^* = n[c, \bar{R}^g(T^*)] \geq n^o = n[w, R^g(T^o)]$ and hence $A^o \subseteq A^*$ (strict inequalites will hold if there is a continuum of sectors). *QED*

Figure 1
The Framework of Analysis



Notation:

GPT:	General Purpose Technology
AS_a:	Application Sector a
z:	"Quality" of the GPT
w:	Market price of the GPT
c:	Marginal cost of GPT
T^a:	Technological level (or "performance") of AS_a (T: vector of T^a's)
π^g, π^a:	Gross rents of the ith sector, i: GPT, AS_a
C^i:	R&D costs of the ith sector, i: GPT, AS_a

Figure 2
Dynamic Reaction Functions

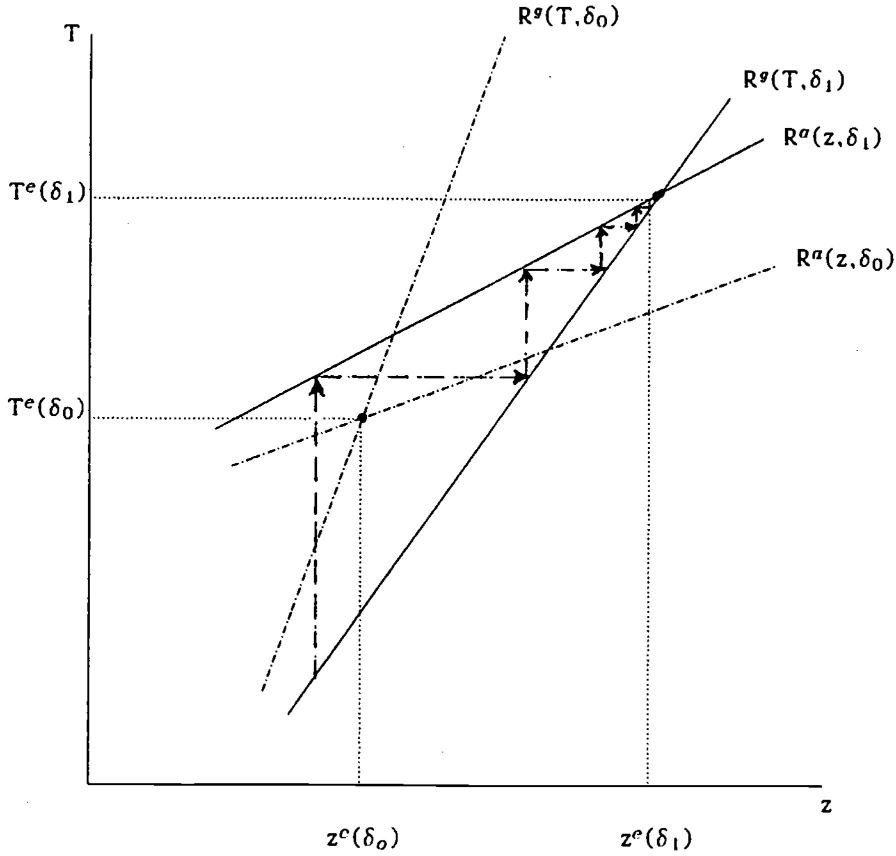
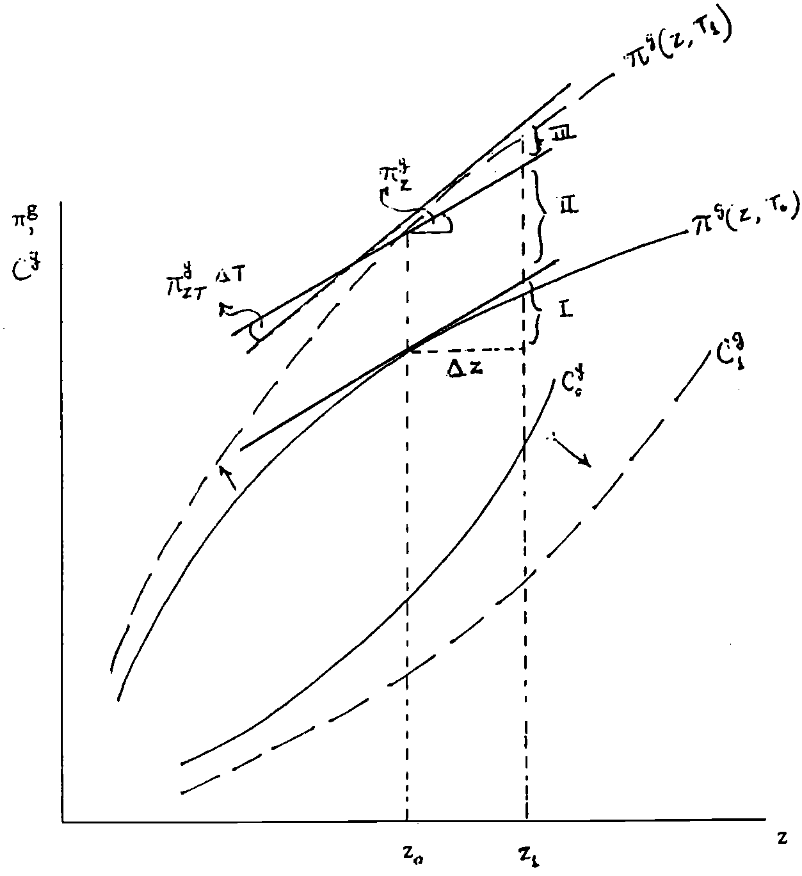


Figure 3

Super-additivity in the Presence of Innovational Complementarities: the GPT Sector



$I = \pi^g_z(z, T_0)\Delta z$

$II = \pi^g_T(z_0, T)\Delta T$

$III = \pi^g_{zT}\Delta z \Delta T$

Table 1

Succesive Generations of a GPT: Actual and Expected

INTEL'S MICROPROCESSOR DYNASTY

Chip	Introduced	
8086/8088	1978/1979	The chips that powered the first IBM PCs and PC clones. They crunch numbers in 16-bit chunks but have limitations in use of computer memory
80286	1982	Speedier than the 8088/8086, the 80286 also enabled computers to run for larger programs. First appeared on the 1984 IBM PC/AT
80386	1985	First Intel 32-bit microprocessor, capable of processing data in 32-bit chunks. Gave PCs power to do bigger jobs, like running networks
80386SX	1988	Lower-priced version of the 80386, aimed at killing off the 80286, which was also produced by Advanced Micro Devices
80486	1989	Intel's "mainframe on a chip." With 1.2 million transistors, it's one of the most complex chips ever made
486SX	1991	The chip aimed at bringing mainframe power to the masses. It will eventually make the 80386 obsolete
586	1992	Expected to have 2 million transistors and at least twice the 80486's performance. Its mission: to compete with RISC chips
686	1993/1994	Just entering the development phase, the 686 is likely to include sound and video-processing features for "multimedia"

DATA: BW

Reproduced from Business Week, April 29, 1991, page 55.