NBER WORKING PAPER SERIES

DIFFUSION OF GENERAL PURPOSE TECHNOLOGIES

Elhanan Helpman
Manuel Trajtenberg

Working Paper 5773

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
September 1996

We are grateful to the National Science Foundation for financial support. 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 5773
September 1996

DIFFUSION OF GENERAL PURPOSE TECHNOLOGIES

ABSTRACT

History and theory alike suggest that General Purpose Technologies (GPT's), such as the steam engine or electricity, may play a key role in economic growth. In a previous paper (Helpman and Trajtenberg, 1994) we incorporated this notion into a Grossman-Helpman growth model, and explored the economy-wide dynamics that a GPT generates. The present paper deals with the diffusion of the GPT over heterogeneous final-good sectors. We show that the gradual adoption of the GPT by each user sector generates a sequence of two-phased cycles, culminating in a "second wave:" after all sectors adopt, they engage anew in R&D, bringing about a spell of sustained growth. We also analyze the welfare implications of the *order* of adoption, by way of numerical simulations. As a "reality check," we sketch the early diffusion of the transistor (the first embodiment of semiconductors, the dominant GPT of our era), and seek to characterize both the early adopters and the laggards in terms of the parameters of the model.

Elhanan Helpman
Department of Economics
Tel Aviv University
Tel Aviv 69978
ISRAEL
and NBER

Manuel Trajtenberg
Department of Economics
Tel Aviv University
Tel Aviv 69978
ISRAEL
and NBER
manuel@ccsg.tau.ac.il

1 Introduction

Both historical evidence and theoretical analysis have brought forth the notion that General Purpose Technologies (GPT's hereafter) may play a key role in economic growth (David, 1991, Bresnahan and Trajtenberg, 1995, Helpman and Trajtenberg, 1994). GPTs are characterized, first, by their pervasiveness, in that they are used as inputs by a wide and ever expanding range of sectors in the economy. As a GPT diffuses it fosters complementary investments and technical change in the user sectors, thereby bringing about sustained and pervasive productivity gains. The steam engine during the first industrial revolution, electricity in the early part of this century and microelectronics in the past two decades are widely thought to have played such a role.

In a previous paper (Helpman and Trajtenberg, 1994) we incorporated the notion of GPTs into a growth model *a la* Judd (1985), and explored the economy-wide dynamics that a GPT may generate. We assumed that advances in the GPT are exogenous, and that they occurred at fixed intervals over time. We found that each time a new GPT appears, it generates a cycle consisting of two distinct phases. The first phase is the "time to sow", in that resources are diverted to the development of complementary inputs that would allow to take advantage of the new GPT. During this initial stage output and productivity experience negative growth, and the real wage stagnates. The "time to reap" comes in the second phase, after enough complementary inputs have been developed, making it worthwhile to switch to manufacturing with the new, more productive GPT. As a consequence there is a spell of growth, with rising output, real wages and profits. As new GPTs appear these cycles are repeated.

In that paper we treated the order of adoption in a crude way. However, the order of adoption, the "lateral" linkages between adopting sectors, and the resulting speed and extent of the diffusion process are all crucial aspects of the dynamic process associated with a GPT, and hence important determinants of the rate of growth induced by GPTs.

Economists have long been aware of the fact that technical advances are surprisingly slow in spreading, and hence that the time profile of the diffusion process is key in determining the realization of the benefits, social and private, from technical change (Griliches, 1957, Mansfield, 1968). However, diffusion studies have typically looked at *individual* innovations, whereas in this paper we seek to analyze the *economy wide* implications of the diffusion of "macro" innovations in the form of GPTs.

The basic framework is that of Helpman and Trajtenberg (1994), except that here we assume that there are a multiplicity of sectors (or final goods), each of them characterized by a set of parameters that account for the fact that some adopt earlier than others, thus generating a diffusion process. The parameter set includes, (*i*) a technology parameter representing the productivity advantage of the new GPT (relative to the old one) for each sector; (*ii*) a "historical stock" parameter indicating the number of components developed for the old GPT; (*iii*) a demand parameter (the spending share of the sector); and (*iv*) an R&D parameter (how expensive it is to develop new, complementary components for the specific sector). The first and second parameters jointly determine the number of new components that have to be developed in order to make it worthwhile to switch to the new GPT; i.e., they determine the height of the "switching threshold."

Assuming an exogenous ordering of sectors, we trace the dynamic trajectory of the economy as one sector after the other adopt the new GPT. Each generates a cycle similar in nature to the economy-wide cycle found in Helpman and Trajtenberg (1994): each cycle is made of two phases, the first depressing real income, the second pushing it up as the sector adopts the new GPT in manufacturing, and hence realizes its benefits in the form of enhanced productivity. As all the sectors adopt and the economy moves towards its steady state, a "second wave" occurs: all sectors engage anew in *R&D*, the benefits of which occur *immediately* (since all sectors are already producing with the new GPT), hence bringing about a spell of sustained and widespread growth.

The picture that emerges is thus as follows. The initial stages in the spread of a new GPT render a process of intermittent growth, as early adopters make the required complementary investments whereas the laggards keep producing with the old GPT. It is only when a critical mass of sectors have already adopted and real income risen enough, that pioneers and laggards alike are induced to make *further* complementary investments that pay off right away, hence generating a spell of sustained growth. We also analyze the welfare implications of the *order* of adoption (by way of numerical simulations), and find that, as intuition would suggest, welfare is higher if the sectors being first to adopt have low switching thresholds, low R&D costs, or high spending shares (the demand parameter).

The setting just described may render a large number of alternative equilibria, since many different sequences of sectors (in terms of order of adoption) may be plausible. Of course, if one were to allow for expectations other than perfect foresight (as assumed here), many more equilibrium trajectories would become feasible. This is by no means a deficiency of the model, but a reflection of the complexity and inherent indeterminacy of the issue at hand. This is clearly manifested in the last section of the paper, where we sketch the early diffusion of the transistor, the first embodiment of the leading GPT of our times, semiconductors.

In particular, we seek to characterize both the early adopters of the transistor (hearing aids and computers), and the laggards (telecommunications and automobiles) in terms of the parameters of the model. Perhaps not surprisingly, virtually *all* parameters mentioned above were particularly advantageous for early adopters (a rare event). Of particular importance seems to have been the plain inadequacy of the incumbent technology for hearing aids and computers vis a vis the new if still rough GPT. As to the laggards, the decisive impediment to early adoption was an extremely large "historical stock", in the form of the sheer size of the system and the complex interdependence of components for telecommunications. For automobiles it referred to the huge stock of manufacturing plants and the disruption that the introduction of the new GPT could bring upon it. Interestingly, the errors of forecast made initially with respect to the laggards were due to the fact that these predictions were based primarily on the expected relative technological advantage of the new GPT, and not on the basis of the height of the threshold that these sectors would have to overcome in order to bring themselves to adopt.

The paper is organized as follows. Section 1 lays out the building blocks of the model. We derive in section 2 the key differential equations driving the system for each sector, at each phase of the cycle. In section 3 we trace the equilibrium trajectories, and in section 4 we discuss the order of adoption, including the welfare analysis through simulations. Section 5 concludes with the historical account of the early diffusion of the transistor.

2 The Building Blocks

There are m sectors. Final output of sector i, Q_i, is produced with an assembly of components $x_i(h)$, $h \in [0, n_i]$, where n_i stands for the number of different components available to the sector. These components have to fulfill two conditions: they have to be compatible with the particular GPT in use by the sector (i.e., components developed for previous GPTs would not work with a subsequent GPT), and they are specific to the sector itself.

2.1 Production

The economy is in a steady state, manufacturing with an old GPT. We begin the analysis at time $t = 0$, when a new more productive GPT arrives unexpectedly. The production function of sector i with the new GPT is

$$Q_i = G_i D_i, \;\; G_i > 1, \tag{1}$$

$$D_i = \left[\int_0^{n_i} x_i(h)^\alpha dh\right]^{1/\alpha}, \quad 0 < \alpha < 1.$$

The elasticity of substitution between any two components is thus $1/(1-\alpha)$, the same in all sectors. But at time zero no components are available for the new GPT. Therefore $n_i(0) = 0$ for all sectors and no production can take place with the new technology.

We can allow a variety of specifications for the old technologies without substantially affecting the analysis. It proves, however, convenient to think about the old GPT as having production functions such as (1) with $G_i = 1$ for every sector. In this event the advantage of the new GPT over the old one is represented by the fact that $G_i > 1$ for the new technology. The larger G_i the bigger the advantage of the new technology for sector i. Next let the number of components available in each sector for the old technology be given. Then manufacturers of final outputs, who are competitive in the product and input markets, can calculate marginal costs for the old technology. They continue to use the old technology as long as these marginal costs are lower than the marginal costs that result from the use of the new technology. But since at time zero there are no components to go with the new technology, marginal costs are necessarily lower for the old technology. Therefore the old technology remains in use at least for some time.

2.2 Pricing

To supply a component for the new technology in sector i, a firm has to invest resources in R&D in order to develop its own unique brand. This invention is protected by a patent, and each such firm engages in monopolistic competition with the other suppliers of components to that sector. All such firms and corresponding components are symmetric, all components can be produced with one unit of labor per unit output, and hence marginal costs equal the wage rate, w, which we normalize as $w = \alpha$. Thus, profit maximizing producers of components, who face demand functions with a constant price elasticity $1/(1-\alpha)$, price them equally;

$$p_i(h) = \frac{w}{\alpha} = 1, \qquad \text{for all } \; h \in [0, n_i] \;\text{ and }\; \text{all } i = 1, .., m. \tag{2}$$

The last equality results from our choice of units (i.e., the wage rate equals α). Under these circumstances all components for the i^{th} sector are employed in equal quantities x_i. As a result profits per new component in sector i are

$$\pi_i = \frac{(1-\alpha)E_i}{n_i}\,, \tag{3}$$

where E_i is spending on good i. This formula applies when the new GPT is used in the sector; otherwise profits equal zero. From (1) and the fact that all components for a given sector are used in equal quantities, it also follows that

$$Q_i = G_i n_i^{\frac{1-\alpha}{\alpha}} X_i, \tag{4}$$

where $X_i = n_i x_i$ is the total employment of inputs. The same type of relationships apply to the old technology.

2.3 Resource constraint

The total amount of labor available in the economy is L, which has to be allocated between the production of components and R&D, where R&D is designed to develop components for the new GPT. Therefore the resource constraint is

$$\sum_{i=1}^{m} X_i^{\circ} + \sum_{i=1}^{m} X_i + \sum_{i=1}^{m} a_i \dot{n}_i = L,$$

where X_i° is the output of old type components for sector i, X_i is the output of new type components for sector i, a_i is the labor requirement per unit invention of new components for sector i and $\dot{n}_i$ is the flow of new components invented for sector i per unit time. We assume for the time being that the R&D coefficients a_i are constant. Since the production of final outputs is competitive and there are constant returns to scale for given components, it follows that $X_i^{\circ} + X_i = E_i$ and therefore that $\sum_{i=1}^{m} X_i^\circ + \sum_{i=1}^{m} X_i = E$, where $E = \sum_{i=1}^{m} E_i$ represents aggregate spending. In this event the resource constraint can be represented as

$$E + \sum_{i=1}^{m} a_i \dot{n}_i = L. \tag{5}$$

In addition, aggregate profits of component manufacturers (new and old) equal $(1-\alpha)E$. Therefore national income is given by

$$Y = \alpha L + (1-\alpha)E, \tag{6}$$

measured in the same units as wages. Divided by α it gives real income in labor units.

In order to take advantage of the new technology, each sector has to develop new components, and enough of them so as to make it worthwhile to switch to the new GPT. We denote by n_{ic} the number of components that have to be developed for sector i in order to induce manufacturers of Q_i to switch to the new technology.[1]

2.4 Profits and the market value of firms

At each point in time consumers maximize a Cobb-Douglas utility function,

$$u = \prod_{i=1}^{m} Q_i^{\beta_i}, \quad \sum \beta_i = 1, \tag{7}$$

implying that $P_i Q_i = E_i = \beta_i E$, where P_i is the price of Q_i. Combined with (3) this relationship implies

$$\pi_i = \frac{(1-\alpha)\beta_i E}{n_i}. \tag{8}$$

The stock market value of each such firm at time t is

$$v_i(t) = \int_t^\infty e^{-[R(\tau)-R(t)]} \pi_i(\tau) d\tau, \tag{9}$$

[1]Unit costs in the production of Q_i are $c_i = X_i/Q_i = 1/G_i n_i^{(1-\alpha)/\alpha}$ with the new technology and $c_i^\circ = X_i^\circ/Q_i^\circ = 1/(n_i^\circ)^{(1-\alpha)/\alpha}$ with the old (see (4)). Therefore the new one becomes profitable whenever $n_i \geq n_{ic} \equiv n_i^\circ/(G_i)^{\alpha/(1-\alpha)}$.

where $R(\tau)$ stands for the discount factor from time τ to zero. We note that this valuation equation implies the no-arbitrage condition

$$\frac{\pi_i}{v_i} + \frac{\dot{v}_i}{v_i} = r, \tag{10}$$

where $r = \dot{R}$ is the instantaneous interest rate.

It pays to develop components for sector i at time t only if the value of a firm is at least as high as the cost of R&D, αa_i. We assume free entry into the R&D business. Therefore whenever new components are developed for sector i the free entry condition implies that

$$v_i = \alpha a_i, \tag{11}$$

whereas no development takes place when $v_i < \alpha a_i$.

2.5 Intertemporal allocation of spending

Consumer preferences at time zero are

$$U(0) = \int_0^\infty e^{-\rho\tau} \log u(\tau) d\tau, \tag{12}$$

where u is given in (7) and ρ is the subjective discount rate. Each consumer maximizes this preference function subject to a budget constraint in which the present value of spending equals the present value of labor income plus the value of firms. The first order conditions of this optimization problem imply

$$e^{-\rho t} E(0) = e^{-R(t)} E(t). \tag{13}$$

Differentiating this condition with respect to time results in the familiar differential equation for aggregate spending

$$\frac{\dot{E}}{E} = r - \rho.$$

Namely, spending grows at a rate that equals the difference between the interest rate and the subjective discount rate. Combined with the no arbitrage condition (10) this implies

$$\frac{\dot{E}}{E} = \frac{\pi_i}{v_i} + \frac{\dot{v}_i}{v_i} - \rho \qquad \text{for every } i. \tag{14}$$

An important property that emerges from (13) is that aggregate spending is a continuous function of time if and only if the discount factor never jumps. We will see, indeed, that the discount factor cannot jump. Therefore spending varies continuously on the equilibrium trajectory, except, of course, at time zero, when the new GPT arrives.

3 Basic Dynamics

A key question in the diffusion of a new general purpose technology concerns the sequence in which the technology is adopted by various sectors. For the moment, however, suppose that this sequence is given. In this event it is convenient to number the sectors according to the order of adoption; i.e., sector $i = 1$ is the first to adopt the new GPT, sector $i = 2$ is the second to adopt it, and so on, until $i = m$. We will discuss the economic factors that determine the order of adoption in section 4.

3.1 Spending

Sector i replaces the old technology with the new one only if the number of components available for the new GPT exceeds n_{ic}. This means that there has to be a time interval in which new components are developed for sector i but in which manufacturers of final output still use the old technology. In this time interval manufacturers of the new components have no buyers and therefore do not produce them, which implies that they have zero profits.[2] In addition, as long as new components are being developed the free entry condition (11) holds, implying that the value of new

[2]Note that this result stems from the fact that there is monopolistic competition in the components sector. As a result component producing firms cannot coordinate their prices in order to induce component buyers to switch to the new technology.

component firms is constant in this industry. But zero profits and a constant value of firms imply (see (14))

$$\frac{\dot{E}}{E} = -\rho. \tag{15}$$

Namely, as long as there is an industry that develops new components but in which the old GPT remains in use, aggregate spending declines at a constant rate that equals the subjective discount rate. It results from the fact that during such time periods the interest rate equals zero (see the no arbitrage condition (10)).

The intuition behind this result is that developers of components who expect to make zero profits for some time prefer to postpone R&D investment as long as the interest rate is positive. Therefore the interest rate has to decline to zero in order to induce them to innovate.

On the other hand, if sector i has already switched to the new technology (i.e., $n_i \geq n_{ic}$), then profits per component are given by (8). In addition, whenever new components for the sector are being developed the free entry condition implies that the value of a component firm is constant. As a result, (8), (11) and (14) imply

$$\frac{\dot{E}}{E} = \frac{(1-\alpha)\beta_i E}{\alpha a_i n_i} - \rho. \tag{16}$$

If more then one sector that has switched to the new technology develops components at the same time, then this equation holds simultaneously for each one of them. Clearly, this can happen only if the ratio $\beta_i / a_i n_i$ is the same in all of them. It is also clear from a comparison of (15) with (16) that it is impossible to have simultaneous development of components in a sector that has not yet switched to the new technology and a sectors that has.

3.2 Research and development

There are three possibilities at a point in time: firms do not invest in R&D, new components are being developed for one sector only, new components are being developed for more than one sector. In the first case the resource constraint (5) implies constant spending; i.e., $E = L$. In the second case it implies

$$a_i \dot{n}_i = L - E, \tag{17}$$

where R&D is taking place only in sector i. This is a differential equation for the number of brands developed in that sector. In the third case the resource constraint implies

$$\sum_{i \in I} a_i \dot{n}_i = L - E, \tag{18}$$

where I is the set of sectors in which R&D takes place. But we have seen that whenever more than one sector innovates, then either all the innovating sectors have not yet switched to the new technology or all of them have (but it is not possible for only some of them to have switched). In the latter case, which is the only case of interest, we must have (see (16))

$$\frac{a_i n_i}{\beta_i} = \gamma_I \quad \text{for all } i \in I. \tag{19}$$

Clearly, the case in which only one sector innovates is a special case in which the set I is a singleton; i.e., $I = \{i\}$.

Why is the case in which innovation takes place in a number of sectors that have not yet switched to the new technology of no particular interest? Because suppose that this is indeed the case. Then if one of these sectors reaches the critical number of components that induces a switch to the new technology and this sector continues to developed components after that point in time, then the interest rate has to be positive after the switch.[3] As a result all other sectors that developed with it, but who

[3]For suppose the interest rate is zero after the switch. Then the no arbitrage condition (10) implies that the value of a components firm that makes positive profits has to decline. On the other hand, as long as there is product development the value of a firm has to be constant and equal to the costs of R&D (see (11)), a contradiction. Therefore the interest rate cannot be zero. The economic reason for the non zero interest rate is that if the interest rate was zero and the value of a profit making firm was expected to remain constant at the level of its R&D costs, then investment in such firms would be infinitely profitable.

have not yet reached the critical number of components that induces a technology switch, regret the early investment in R&D (they would be better off postponing the investment in product development). But this type of regret cannot happen on an equilibrium trajectory. Therefore either all of them reach the critical number of components simultaneously or the sector that reaches this point first stops further product development. Both are knife edge cases that we assume away.

3.3 Continuous discounting

Consider a point in time t at which some R&D takes place. It is apparent from the valuation of firms in (9) that v_i is a continuous function of time. Therefore postponing at t the development of a product by a negligible amount of time has a negligible effect on the reward of doing R&D. At the same time the postponement of product development by a negligible amount of time has a negligible effect on R&D costs if and only if the discount factor $R(\cdot)$ is continuous to the right of t. For example, if the discount factor is discontinuous to the right of t and $R(t^+) > R(t)$, it is preferable to invest in product development a moment after t, and if $R(t^+) < R(t)$ than it is strictly preferable to invest in R&D at t as compared with a moment later. A similar argument can be made with respect to discontinuities of $R(\cdot)$ to the left of t. It follows that there can be no investment in product development over an entire interval of time unless the discount factor is continuous on this time interval. We seek equilibrium trajectories on which there is research and development from the moment of arrival of the new GPT and until the economy reaches a new steady sate.[4] On all such trajectories the discount factor has to be a continuous function of time. It then follows from the consumers' first order condition (13) that on all such trajectories aggregate spending is a continuous function of time.

[4]It is very unlikely that periods in which there is no product development will follow and precede periods with product development. This point will become apparent when we discuss the determination of the initial value of spending.

3.4 Sectoral wave

As indicated at the beginning of this section, for the time being we take as given the order of adoption of the new technology. In this event we may think of a typical time interval over which firms develop components for sector i, which follows a time interval over which R&D was directed towards product development in the preceding sectors. Suppose that the development of components for sector i begins at time T_i and continues until T_{i+1}, with no components for other sectors being developed in the meantime. Also suppose that in the time interval $[T_i, T_{i+1})$ manufacturers of final output Q_i have switched at some point to the new technology. Then we can divide this time interval into two phases: in phase 1 components for i are being developed while the old technology is in use while in phase 2 components for i are being developed while the new technology is in use. The switch from phase 1 to 2 occurs when n_{ic} components become available.

In phase 1 aggregate spending evolves according to (15) while the number of components evolves according to (17). The phase begins with $n_i = 0$ and ends with $n_i = n_{ic}$. It is characterized by declining spending and a rising number of components, as depicted in Figure 1. The initial level of spending $E(T_i)$ equals the level of spending at time T_i^-, because spending is continuous in time. Given this initial level of spending our differential equations determine the level of spending over the entire phase, and in particular $E(T_{ic})$, where T_{ic} is the point in time at which phase 1 ends. The phase ends when the number of available components equals n_{ic} and manufactures of final output switch to the new technology.

By continuity, phase 2 begins with the spending level $E(T_{ic})$ and the number of components n_{ic}. Starting with these initial conditions spending in phase 2 evolves according to (16) while (17) continues to be the differential equation for the number of components. The number of components continues to rise, but spending may rise or decline. Figure 1 depicts a phase-2 trajectory on which spending rises. Importantly, the end of this phase determines a number of components $n_i(T_{i+1})$ and a level of spending $E(T_{i+1})$. This spending level serves as the initial spending level for the dynamics that follow.

The two phases combined generate a sectoral wave of the type depicted in Figure 1. It remains to be seen how such sectoral waves fit in into equilibrium trajectories.

4 Equilibrium Trajectories

Building blocks for an analysis of economic dynamics have been developed in the previous section. We now use them to construct equilibrium trajectories. For this purpose first note that in economies of this type perfect foresight leads to a steady state in the long run. We therefore seek trajectories that consist of two-phase sectoral waves that converge to a steady state. To this end we first describe the steady state.

4.1 Steady state

In a long-run steady state there is no product development. Therefore by the resource constraint (5) aggregate spending equals the labor force:[5]

$$\tilde{E} = L. \tag{20}$$

Under these circumstances (13) implies that the interest rate equals the subjective discount rate. We consider economies in which all m sectors adopt the new technology. Therefore (8), (9) and (20) together with the fact that the interest rate equals ρ imply

$$\tilde{v}_i = \frac{(1-\alpha)\beta_i L}{\rho \tilde{n}_i} \qquad \text{for all } i = 1, 2, ..., m. \tag{21}$$

Clearly, this value of a firm cannot exceed αa_i, because if it were it would be profitable to further invest in product development. It follows that $\tilde{n}_i \geq (1-\alpha)\beta_i L/\rho\alpha a_i$. On the other hand, if $\tilde{n}_i > (1-\alpha)\beta_i L/\rho\alpha a_i$ then it cannot be that this number of components has been attained by investment in R&D in the last phase of the

[5]We use tildes to denote steady state values.

equilibrium trajectory, just prior to its reaching the steady state. To see why, observe that as long as there is R&D the resource constraint implies that aggregate spending is smaller than L. Therefore aggregate spending is rising during this last phase in order to reach $\tilde{E} = L$ while the differential equations that drive the system consist of (16) for the spending level (for $i \in I$) and (18) for the number of components, with the side condition (19). Since E is rising, it follows from (16) that $n_i < (1-\alpha)\beta_i L/\rho\alpha a_i$ during this phase, which proves our claim. But if components for sector i are not developed during this last phase, then their development ceased at some earlier point in time, say at T_{i+1}. If so, then (8), (9) and (13) imply that at that point in time the value of a components firm in industry i was $(1-\alpha)\beta_i E(T_{i+1})/\rho\tilde{n}_i$ and that this value equaled product development costs αa_i, implying $\tilde{n}_i = (1-\alpha)\beta_i E(T_{i+1})/\rho\alpha a_i$. Because there was product development at time T_{i+1}, however, aggregated spending was smaller than L (due to the resource constraint), implying $\tilde{n}_i < (1-\alpha)\beta_i L/\rho\alpha a_i$; a contradiction. We therefore conclude that

$$\tilde{n}_i = \frac{(1-\alpha)\beta_i L}{\rho\alpha a_i} \qquad \text{for all } i = 1, 2, ..., m, \tag{22}$$

and that there is product development in all sectors when the economy converges to the steady state.

4.2 Second wave

Our analysis identified a number of features in and around the steady state. Importantly, every sector has to engage in product development when the economy converges to the steady state. This means that all sectors that stopped developing new components at some earlier point in time are bound to renew product development during this last phase. As a result, *there has to be a second wave of product development for each sector, except for the last one.*

The fact that each sector has to invest in R&D when the economy approaches the steady state proves that there has to be a second wave of product development, but it does not exclude the possibility that some sectors will experience more than two. The dynamics of, second, third, etc., waves of product development are described by (16) for spending and by (18) for the number of components, with the side condition (19), where I is the set of sectors that jointly engage in R&D. The same system can be represented in an alternative form that suppresses the side condition;

$$\frac{\dot{E}}{E} = \frac{(1-\alpha)E}{\alpha\gamma_I} - \rho, \tag{23}$$

$$\dot{\gamma}_I \sum_{i \in I} \beta_i = L - E. \tag{24}$$

During the convergence to a steady state this system applies with $I = \{1, 2, ..., m\}$, and therefore $\sum_{i \in I} \beta_i = 1$.

Figure 2 describes the unique saddle path of (23) and (24) that leads to a steady state. On this trajectory spending rises and so does the number of components in each sector. The equilibrium trajectory hits this saddle path at some point in time and coincides with it from this point on. In order for all sectors to engage in R&D simultaneously, however, they have to begin R&D activities in a suitable order, following the preceding sectoral waves. This order is endogenous, given the order in which the sectoral waves evolve.

To determine the order in which sectors join the last phase of product development, suppose that sector i ended the most recent period of product development at time $\hat{T}_i$ with $\hat{n}_i = n_i\left(\hat{T}_i\right)$ components. As a result sector i does not engage in product development from $\hat{T}_i^+$ until it joins the last phase of development. Given the values of $\hat{n}_i$ for all $i = 1, 2, ..., m-1$, we can identify the order and timing by means of the following

Last Phase Procedure

1. Start from the steady state.

2. Run backwards the differential equations (23) and (24) for $I = \{1, 2, ..., m\}$ and calculate $E(t)$ and $n_i(t)$. The number of components declines for each sector in this backward calculation.

3. Let i_{m-1} be the first sector for whom $n_i(t) = \hat{n}_i$, and let this happen at time $\bar{T}_{i_{m-1}}$. Then this is the last sector to join the last wave of development.

4. Starting with the spending level and the number of products that were calculated in the previous step for $t = \bar{T}_{i_{m-1}}$, repeat step 2 with $I = \{1, 2, ..., m-1\} - \{i_{m-1}\}$ (i.e., by dropping sector i_{m-1} from the set of sectors that are active in R&D). Stop this calculation at time $\bar{T}_{i_{m-2}}$ when a second sector, i_{m-2}, satisfies $n_i(t) = \hat{n}_i$. Sector i_{m-2} is the second to the last to join the last wave of development.

5. Proceed in similar fashion to identify the order of all remaining sectors.

At the end of the process we obtain a perturbation $(i_1, i_2, ..., i_{m-1})$ of $(1, 2, ..., m-1)$, which represents the order in which the sectors begin the last phase of product development; sector i_1 is the first to begin this last phase while i_{m-1} is the last. We denote by $\iota(i)$ the order of sector i in this perturbation. Namely, $\iota(i) = j$ if and only if $i_j = i$. This order depends, however, on the values of $\hat{n}_i$, $i = 1, 2, ..., m-1$, which are also endogenous. It is therefore necessary to examine the entire trajectory in order to identify the equilibrium values of $\hat{n}_i$.

4.3 Entire trajectory

Our first task in describing the equilibrium trajectory is to characterize the equilibrium values of $\hat{n}_i$. For simplicity we restrict this discussion to trajectories on which each industry experiences a sectoral wave and renews R&D only during the last phase of product development. Namely, we abstract from situations in which an industry experiences a sectoral wave, develops new products at some later stage (or stages) as well, and finally joins the last phase of product development. Such equilibrium trajectories cannot be excluded, however, and we will discuss them briefly later on.

So suppose that we run (23) and (24) backwards, using the five steps of the Last Phase Procedure. At the end of the process we obtain an order in which sectors join the last phase of product development, i_j, $j = 1, 2, ..., m-1$, and the points in time at which they join in, $\bar{T}_{i_j}$. Now proceed to the

Fixed Point Procedure

1. Start at $\bar{T}_1$ with $E\left(\bar{T}_1\right)$ and $n_m\left(\bar{T}_1\right)$. At this point in time the last sector, m, completes its sectoral wave.

2. Run backwards the differential equations (16) and (17) for $i = m$, until $n_m(t) = n_{mc}$. This identifies phase 2 of the sectoral wave in m. In particular, it identifies the point in time at which phase 2 begins, T_{mc}, and the spending level $E(T_{mc})$. Naturally, $n_m(T_{mc}) = n_{mc}$.

3. Starting with $E(T_{mc})$ and $n_m(T_{mc}) = n_{mc}$ run backwards the differential equations (15) and (17) for $i = m$ until $n_m(t) = 0$. This identifies phase 1 of the sectoral wave in m. In particular, it identifies the point in time at which phase 1 begins, T_m, and the spending level $E(T_m)$. This point in time also represents the end of phase 2 in sector $m-1$.

4. Sector $m-1$ ends its sectoral wave with $n_{m-1}(T_m) = \hat{n}_{m-1}$ components and it joins the last phase of product development at time $\bar{T}_{\iota(m-1)}$. Between time T_m and $\bar{T}_{\iota(m-1)}$ it does not invest in R&D. It therefore follows from the valuation of firms (9) and the free entry condition (11) that $\hat{n}_{m-1}$ has to satisfy[6]

$$\hat{n}_{m-1} = \frac{\beta(1-\alpha)E(T_m)}{\rho\alpha a_{m-1}} \frac{1 - e^{-\left[\bar{T}_{\iota(m-1)} - T_m\right]}}{1 - \frac{E(T_m)}{E\left(\bar{T}_{\iota(m-1)}\right)} e^{-\left[\bar{T}_{\iota(m-1)} - T_m\right]}}. \tag{25}$$

[6]The valuation equation (9) implies that for every t_1 and $t_2 > t_1$

$$v_i(t_1) = \int_{t_1}^{t_2} e^{-[R(t) - R(t_1)]}\pi_i(t)\,dt + e^{-[R(t_2) - R(t_1)]}v_i(t_2).$$

In our case $t_1 = T_m$ and $t_2 = \bar{T}_{\iota(m-1)}$, the free entry condition (11) holds at the two end points, and profits satisfy (8) at each point in time. Combined with the consumer's first order condition, (13), these conditions and the valuation equation imply (25).

If this condition does not hold, it means that the Last Phase Procedure did not use the equilibrium value of $\hat{n}_{m-1}$.

5. Starting with $E(T_m)$ and $n_{m-1}(T_m) = \hat{n}_{m-1}$, repeat steps 2-4 (with suitably modified indexes) for sector $i = m-1$, then for $i = m-2$, etc., until $i = 1$. This traces out the remaining parts of the equilibrium trajectory.

It is clear from these procedure that they can be used to find the equilibrium values of $\hat{n}_i$ as a fixed point of a mapping. Given a vector $(\hat{n}_1, \hat{n}_2, ..., \hat{n}_{m-1})$ our Last Phase and Fixed Point procedures map this vector into another vector by means of (25). A fixed point of this mapping identifies the equilibrium values.

Figure 3 depicts an equilibrium trajectory for a three-sector economy, combining the features of the sectoral wave in Figure 1 with the convergence to a steady state in Figure 2. The aggregate number of components is measured on the horizontal axis. Product development begins in sector 1. Its first phase lasts until n_{1c} components have been developed, whereupon its second phase begins, after manufacturers of final output switch to the new technology. The second phase ends when there are $\hat{n}_1$ components in sector 1. Then sector 2 begins the first phase of its sectoral wave. This one lasts until there are n_{2c} components to go with the new technology in sector 2. And so on. The third sector completes its sectoral wave when there are $n_{S1} = \sum_{i=1}^{3} \hat{n}_i$ components. At this point sector i_1 (which is either sector 1 or 2) joins the last phase of product development. Sectors 3 and i_1 continue to invest in R&D until the number of components equals n_{S2}, at which point the remaining sector i_2 also joins the last phase of development. From this point on all three sectors engage in R&D until the economy approaches the steady state, in which product development ceases.

Figure 3 describes an equilibrium trajectory on which the second wave is also the last phase of product development, which is the case on which our analytical part has focused. To ensure that this is indeed an equilibrium, it also has to satisfy $v_i(t) \leq \alpha a_i$ at each point in time for every sector. Naturally, during periods of product development for sector i this holds as an equality. If we calculate a candidate equilibrium by means of our procedures and it turns out that this inequality does not hold for some sector j in some time interval, it means that the second wave is not the last one for this sector. It is then necessary to search for equilibrium trajectories in which sector j starts a second wave of product development before the last, sometime during phase 2 of a sectoral wave in some other sector with a higher index. As a consequence sector j experiences more than two waves of product development. We will not elaborate a procedure for calculating this type of trajectories.

4.4 Real wages and real income

By our normalization, the wage rate equals α at each point in time while income, Y, equals $\alpha L + (1-\alpha)E$ (see (6)). In order to obtain the real wage rate and real income we need to deflate these quantities by a suitable price index. From individual preferences (7) the price index of real consumption can be represented by

$$P = P_0 \prod_{i=1}^{m} P_i^{\beta_i},$$

where P_0 is some constant and P_i is the price of final output in sector i. On the other hand, it follows from (4) and the definition of n_{ic} (see footnote 2.3) that the price of final output in i, which equals unit costs, is given by

$$P_i = \begin{cases} G_i^{-1} n_{ic}^{-(1-\alpha)/\alpha} & \text{for } n_i \leq n_{ic}, \\ G_i^{-1} n_i^{-(1-\alpha)/\alpha} & \text{for } n_i \geq n_{ic}. \end{cases}$$

By a suitable choice of P_0 we can therefore represent the price index of consumption as

$$P = \prod_{i=1}^{m} \phi_i(n_i)^{\beta_i}, \tag{26}$$

$$\phi_i(n_i) = \begin{cases} n_{ic}^{-(1-\alpha)/\alpha} & \text{for } n_i \leq n_{ic}, \\ n_i^{-(1-\alpha)/\alpha} & \text{for } n_i \geq n_{ic}. \end{cases}$$

The real wage rate α/P on the trajectory from Figure 3 is depicted in Figure 4. It remains constant at its initial level during the first phase of the first sectoral wave and it rises steadily during phase 2 of this sectoral wave. This stems from the fact that as long as sector 1 has not switched to the new technology the purchasing power of wages remains constant. But once it has switched, the price of good 1 declines as long as additional components are developed for sector 1. As a result real wages rise. A similar pattern of real wages evolves in each subsequent sectoral wave. In the final phase, when all sectors have switched to the new technology, real wages rise steadily as long as R&D is positive. Real wages stabilize when the economy settles down on the steady state. Our model predicts, therefore, periods of rising real wages, followed by periods of stagnation, as the new technology diffuses across sectors.

Figure 5 depicts the pattern of real income, $Y/P = [\alpha L + (1-\alpha)E]/P$, measured in real consumption units. The arrival of the new GPT leads to a drop in real income, as resources shift from manufacturing to R&D. This is similar to the result in Helpman and Trajtenberg (1995). From this point on real income declines continuously during phase 1 of the first sectoral wave and rises during phase 2. The same pattern is exhibited in subsequent sectoral waves, until the last phase in which real income rises until the economy arrives at the steady state. During the first phase of each sectoral wave prices of final goods are constant but demand for them declines. This falling demand, which is accompanied by rising investment in R&D, reduces real income, because due to the monopolistic price distortion in the production of intermediates the employment of resources in manufacturing is too small to begin with. In phase 2 of a sectoral wave aggregate demand for final goods rises and one of the prices declines. On account of both these trends real income rises. Since in the steady state real income is higher than in the initial steady state, it follows that the economy experiences average productivity growth, but that this positive average growth is driven by cycles of declining and rising productivity levels as the use of the new technology spreads across sectors.

5 Order of Adoption

Our analysis of equilibrium trajectories assumed a known order in which the new technology spreads across sectors. Given this order, there are sectoral waves and a final phase of across-the-board product development.

In this section we raise two questions. First, what features determine the order of adoption? And second, if there are multiple equilibria that differ in this order, how are they ranked according to relative efficiency?

5.1 Multiple Equilibria

The first thing to note is that there are multiple equilibrium trajectories in economies of this sort. To see why, observe that the profitability of developing a component for a given sector depends on expectations as to whether others will also develop new components for this industry. If one potential component developer expects that others will not invest, he understands that the sector will not adopt the new technology even if he does develop his particular brand. As a result, he will not be able to sell any output and therefore will not be able to cover R&D costs. Consequently, he does not develop the component.

This argument makes clear a critical point; R&D requires coordination of expectations among potential developers. If all believe that there will be sufficient development to induce the manufacturers of final output to switch to the new technology, then they invest in R&D and thereby justify these expectations. If, however, they do not believe that there will be enough product development, then none of them invests in R&D and these expectations are also justified. The trajectories described in the previous section assume optimistic expectations that are justified in retrospect.

Our discussion suggests that, in addition to an equilibrium in which all sectors end up adopting the new technology, there also may exist equilibria in which one or more of the industries continue to use the old technology and no one in these sectors invests in R&D. Another point that becomes clear from this discussion is that not only expectations about whether there will be enough product development in a particular sector matters; a potential investor has to form also expectations about the timing of product development by his rivals. If, for example, a company expects that all the others will start product development in the very distant future, and that as a result manufacturers of final output will switch to the new technology in a very distant point in time, then this company finds it unprofitable to invest today. As a result it delays product development.

Figure 6 presents simulated equilibrium trajectories for a two-sector economy in which the only difference between the sectors is the critical number of components that need to be developed in order to induce manufacturers of final output to switch to the new technology. In panel (a) the new GPT is first adopted by the industry that needs more components while in panel (b) it is first adopted by the industry that needs fewer components. Importantly, both are equilibrium trajectories on which investors have perfect foresight. They illustrate the point made above; i.e., that there typically exist multiple equilibria that differ in the order of adoption. This two-sector example has two additional equilibria: one in which only the sector with the smaller critical number n_{ic} develops components for the new GPT (and the other sector remains with the old technology), and one in which only the sector with the larger critical number n_{ic} develops components for the new GPT. Evidently, multiplicity of equilibria is a generic property of this type of economies.

5.2 Efficiency

Which order of adoption is preferable? To answer this question we need to compare the values of $U(0)$ (given in (12)) that result from alternative trajectories. Let $U^j(0)$ be the utility level that obtains from some trajectory j, which is characterized by a particular order of adoption. Then it follows from (7) and (12) that the difference in utility levels from trajectories j and k is given by

$$\begin{aligned} U^j(0) - U^k(0) &= \int_0^{+\infty} e^{-\rho t} \left[\log E^j(t) - \log P^j(t)\right] dt \\ &\quad - \int_0^{+\infty} e^{-\rho t} \left[\log E^k(t) - \log P^k(t)\right] dt, \end{aligned}$$

where $E^j(t)$ is the spending level at time t on trajectory j and $P^j(t)$ is the value of the price index (26) at time t on trajectory j. Once the suitable trajectories are known, we can calculate these welfare differences.

For the two trajectories depicted in Figure 6 we find that welfare is higher in panel (b), when the sector that needs fewer components in order to switch to the new technology invests first. This accords well with intuition, because the sector with the larger critical value n_{ic} has a disadvantage over the other one, in the sense that it needs to develop more components in order to make the new technology viable. Therefore it is better to let it develop last, when the price of consumption is low, and therefore when the alternative cost of R&D is lower.

Although this welfare ranking makes sense and we found it in all the simulations, it is important to note that it is not entirely trivial. To see why, consider Figure 7 in which we plot real consumption E/P for trajectories with alternative orders of entry. Evidently, when the disadvantaged sector invests first real consumption is slightly higher for some time, lower for a span of time afterwards, and it becomes higher again. An examination of the figure suggests that the welfare comparison may depend on the subjective discount rate. But we were not able to find a case in which early entry of the disadvantaged sectors is more desirable.

The association between welfare and the order of adoption is also clear-cut when sectors differ in spending shares. Intuitively, one would expect higher welfare when the sector with the larger spending share invests first, and this has turned out to be the case in all the simulations. Finally we examined differences in labor inputs per unit R&D. According to intuition, welfare should be higher when the sector with the lower R&D costs invests first, and this is indeed confirmed by our simulations.

These examples demonstrate that the desired order of adoption depends in a clear way on the underlying parameters, when we examine one parameter at a time. When there are conflicting differences in more than one parameter, however, then the desired order of adoption becomes a complicated function of the economy's structure. Thus, for example, if the sector with the larger spending share also requires more inputs per unit R&D, then it is not possible to provide a simple characterization of the required strength of these differences that tilts the welfare advantage in favor of one sector rather than the other. As a result there do not exist simple criteria upon which to judge which order of adoption is preferable.

Moreover, given that there are multiple equilibria, we may observe efficient as well as inefficient trajectories; there does not appear to exist a market mechanism that will coordinate expectations upon which the order of investment is determined so as to ensure the emergence of efficient outcomes. As a result, in a multi-sector economy many outcomes are equally plausible. In such circumstances there is plenty of room for historical accident and path dependence.

5.3 Limitations

Two limitations of the model need to be brought to the forefront. The first is that we assume that once a new GPT appears, its relative productivity level G_i remains constant over time. Of course, the model does allow for further improvements, but only trivially, in the form of the appearance of a new GPT. This is clearly in sharp contrast to what we observe in reality: these technologies keep evolving and improving, within the span of what we would regard as a given GPT. Moreover, the cumulative effect of these improvements can be very substantial, such that by the time a new GPT appears, it has to compete with a much better incumbent. We have not incorporated a changing GPT simply because the model is already too complex to handle this additional feature. It would seem though that a changing GPT would have two opposing effects: within the span of a given GPT, diffusion and growth would certainly accelerate. However, and for the same underlying reasons, the adoption of new GPTs may be delayed.

The second limitation refers to the absence in the model of lateral linkages (or spillovers) between adopting sectors. In fact though it is quite likely that sectors that are "close" to each other in the relevant metric,[7] will benefit from positive externalities from neighboring sectors. For example, there may be a learning effect of the following form,

$$a_i = \phi_i(n_{i-1}), \;\; \phi' < 0.$$

That is, the more components for the new GPT developed by the "previous" sector, the lower the development costs of the sector next in line. This constitutes then a "supply push" mechanism, that may play an important role in the diffusion process. We have explored extensions of the model that incorporate this feature, but in order to make it operative we had to resort to other limiting (and equally unsatisfactory) assumptions, such as the absence of a "second wave." It would certainly be interesting to try to pursue extensions of the model in this direction.

6 "Reality Check"

In this section we sketch the initial diffusion stages of the paramount GPT of our times, semiconductors, and seek to characterize both the leading and the laggard user sectors, in terms of the model developed above.

In December 1947, John Barteen and Walter Brattain – of the Bell Labs – demonstrated the feasibility of the point contact transistor.[8] In January 1948 they applied

[7]Such as "technological proximity" to the new GPT. An example would be the extent to which potential user sectors relied on vacuum tubes prior to the appearance of the transistor (see section 6).

[8]William Shockley was also a key player in the team that developed the transistor, and indeed the three of them received the Nobel Prize in Physics for their achievements in this field. In 1947-

for a patent, and in July of that year Bell Labs announced it to the world. A new GPT, semiconductors, was born. Almost half a century later we are still in the midst of the semiconductor era, which is leading us into the next stage in its meteoric rise, the so called "information age." One of the idiosyncratic features of this GPT is that it keeps getting smaller as it gets more powerful. Thus, paradoxically, transistors, integrated circuits, microprocessors and similar microelectronic components tend to be further removed from our sight and perception as they become ever more pervasive. It is therefore easy to overlook the fact that semiconductors are very much at the heart of the current wave of feverish innovation revolving around the Internet, and around the much heralded convergence of telecommunications and computers.

The first decade in the evolution of semiconductor technology was characterized by rapid advances in the manufacturing of transistors (which made them a viable commercial component), by the gradual switch from germanium to silicon, and by vast improvements in the performance characteristics of the transistors themselves. The next quantum leap occurred in 1959, with the invention of the integrated circuit (by Jack Kilby, of Texas Instruments), which allowed for the packing of a large number of electronic components in a single, tiny chip. A decade later it was the turn of VLSI (very large scale integration), and later on that of the microprocessor, the heart of personal computers, developed at Intel.

Even though all of these developments are part and parcel of the same GPT, in terms of our model each can be thought of as a further upward jump in G_i,[9] that would trigger new waves of complementary developments and hence new dynamic trajectories. We concentrate here on the first decade of the new GPT, that of the transistor, and explore which sectors were the first to adopt it, and how they can be

48 Shockley was pursuing a different track, that proved to be superior to the initial point contact transistor: that of the junction transistor, which he domonstrated in 1951.

[9]Note that these further jumps have to be totally unexpected in order to qualify as such in the context of our model. Of course, the G_i's keep advancing over time within these sub-periods (or "generations"), but our model ignores that.

characterized in terms of the parameters of our model. Similarly, we look into those sectors which were thought at the time as natural candidates for rapid adoption, but turned out in fact to be laggards. The early user sectors were hearing aids and computers.[10] The prominent laggards were telecommunications and automobiles.[11]

6.1 The Early Adopters

Hearing Aids

The first sector to adopt the transistor almost immediately after it went into commercial production (in the early 1950s) was hearing aids. In fact, by 1953 more than 15 manufacturers of hearing aids were buying transistors from Raytheon, then the leading producer of point contact transistors. What made the transistor so attractive for hearing aids at that early stage was first and foremost a large relative technological advantage vis a vis the incumbent technology, vacuum tubes (to be referred also as "valves").[12] First, transistors were much smaller and lighter than the (miniature) valves they replaced: the volume of a transistor-based hearing aid was about 45 cc versus 130 cc for a valve-based device.[13] Second, the power requirements were significantly lower, and hence the batteries that operated the system lasted much longer. In fact, the battery life increased *8 times*, and the annual battery replacement cost

[10]By "sector" we mean here a particular end-product or industry that uses semiconductors as components. It may be literaly a whole sector (in the SIC sense), or more likely, a particular product family (such as computers).

[11]The discussion here draws heavily from Braun and MacDonald (1982), hereforth just B&M.

[12]It is important to emphasize that what counts is indeed the relative technological lead in the context of the particular sector. That is, it may be that G_i is large for a given sector not because the new GPT (which is common to all sectors) is technologically very advanced at that stage, but because the previous GPT is particularly inadequate for the needs of that sectors. This was clearly, and strikinly so, the case for early computers (see below).

[13]It is interesting to note that valves had been getting smaller for over a decade prior to the arrival of the transistor, and were already surprinsingly small. Still, the transistor was an order of magnitude smaller.

went down from $40 to just $3. Third, transistors acted instantly, since there was no filament to heat. In terms of our model these relative advantages would be captured by a large G_i.

But there were also distinct disadvantages: early transistors were noisier than valves, more restricted in their frequency performance, and more liable to damage by power surges and high temperatures. And there was a large price differential: early transistors were much more expensive than valves (8 times more), and hence the price of transistor-based hearing aids was 3 - 4 times as expensive as conventional ones: $150-200 (in 1953) versus $50 for hearing aids using subminiature valves. Nevertheless, hearing impaired people were willing to pay the much higher price, which implies that they put a high value on the relative advantages of the new technology. However, note that $200 was not an exorbitant price in 1953 (it would be $1,140 in 1995 dollars), and hence that this was a very *affordable* early application. Of course, this was a very small market, made up of a well defined group of users, that could easily evaluate the price/performance of a transistor-based system vis a vis the incumbent technology.

Computers

One of the most attractive initial uses of transistors was in early computers. Electronic computers were first built just before the transistor came into being (in the late forties). These were electronic monsters, containing thousands of valves, occupying large areas, requiring vast power consumption, and facing tremendous problems because of valve failure, overheating, etc. The transistor was thus ideally positioned to replace the valves and bring about dramatic improvements in those dimensions, as well as in speed (primarily because of the much shorter circuitry). Already in 1955 IBM brought out a transistor-based computer that replaced over 1,000 valves with 2,200 transistors; the new model was significantly smaller, there was little need for cooling, and power consumption was reduced by 95%. The series IBM 7000 quickly followed, using high speed transistors (the IBM 7090 for example, ran about 5 times as fast as the valve-based 709; see Williams, 1985). Transistors rapidly and completely displaced vacuum tubes in computers, and the computer sector established itself as the leading user of semiconductors (outside the Military) for decades to come.[14]

As with hearing aids, early users displayed a high willingness to pay for the much faster, transistor-based computers, and there were relatively few of them. However, the underlying market could not have been more different. Most of the early computers were sold (or leased) to US Government agencies, particularly the Military and defense contractors (having at their disposal generous budgets), and they cost several hundred thousand dollars.[15] The number of computers sold in the early 1950s was very small (a total of 250 computers by 1955), and overall demand was estimated at first to be very restricted.[16] However, dramatic advances in both hardware and software and concomitant reductions in "real" prices brought about an enormous expansion in the range of scientific and commercial applications, and hence in the size of the market. Computers are thus one of the first, and certainly the most striking example of the power of the new GPT, not just as a better substitute for older technologies, but as the key to a whole range of new applications and uses.

Other early adopters

Two additional sectors have to be mentioned in this context: first, the Military played an extremely important role almost from the beginning, both as the largest single user of transistors (since the mid 1950s), and in fostering research and innovation in this field. However, we focus here on *commercial* sectors and hence we do

[14]Computers accounted for 31% of all semiconductors outside the government (and hence outside the military) in 1963, 65% in 1978, and 55% in 1983 (B&M, p. 80, and p. 144).

[15]In 1955 the average price of CPUs of Computer Systems was over $600,000, and it represented about 80% of the total cost of the system (see Flamm, 1987, Table A.1). Nominal prices declined significantly over the course of the following decade (the average CPU cost $88,000 in 1965), and of course quality adjusted prices dropped at a much faster rate.

[16]A survey by the Department of Commerce conducted in the late 1940's estimated that about 100 mainframe computers would satisfy the entire needs of the nation (see B&M, p. 69).

not expand on it. Second, radios were also an important early adopter: Texas Instruments introduced the transistor radio in October 1954, in 1955 Raytheon brought out its own model (selling for $80), and shortly afterwards car radios were also built with transistors. Transistor radios rapidly gained wide acceptance, and in fact they became the first transistor-based *mass* consumer product. The lack of more detailed information about this sector prevents us from further elaborating on it.

The following figures give a sense of the relative weight of the sectors reviewed above, as well as of other adopting sectors in the market for transistors, by the end of the decade under examination (adapted from B&M, p. 80, Table 7.5):

Value of US Transistors by Usage, 1963 ($M)

Military	119
Computers	42
Radios	33
Communications	16
Test and Measuring	12
Controls	12
Other Industries	12
Hearing Aids and Organs	7

6.2 The Laggards[17]

Telecommunications

Writing in the late 1990s, it is ironic to talk about the telecommunication sector ("telecom" in short) as a "laggard" in the adoption of semiconductor technology. After all, we have seen in the past two decades the "digitization" of the telecom sector, the successful marriage of computers and telecommunications, and the proliferation

[17]Another prominent laggard was Television sets. It seems that the dominant reason for the slow incorporation of transistors in TV sets was purely technological, but this requires further investigation.

of new modes of telecommunications based on both. Moreover, these developments constitute one of the outmost expressions of the power of semiconductors, of their extremely wide reach, and of their potential for bringing forth deep and pervasive changes in the economy and society.

Yet, the telecom sector turned out in fact to be a "laggard" in the early era of the transistor: adoption of semiconductor technology was extremely slow, and did not have a significant impact on the sector as a whole until much later. This was so despite the fact that the drive towards the development of semiconductors at Bell Labs was motivated to a significant extent by the perception that the telecom sector would greatly benefit from the incipient technology. Consequently, telecom was widely seen at the time as the natural candidate for early and widespread adoption of the transistor. In that sense this is as good an example as any of wrong technological expectations.

The flip side of the potential for great benefits is the potential for extensive disruption. Telecom is not "one thing" but a huge, intricate, and delicate system, that poses strict demands on the technologies from which it feeds (such as compatibility, reliability, durability, etc.). Moreover, technology is embedded in a massive stock of capital, and hence any change in the underlying GPT requires large capital outlays, and concomitant adjustment costs. The sheer magnitude of these outlays, coupled with the rapid pace of change in semiconductor technology and the difficulties in forecasting future developments, made it extremely difficult to undertake decisions about adoption.[18] The result was that the adoption of semiconductor technology in the telecom sector proceeded at a much slower pace than anticipated. Summarizing the difficulties of adoption, Braun and MacDonald write,

"It is because the adoption of this technological innovation affects so

[18]See for example Braun and MacDonald's detailed account of the mistakes made by the british Post Office (now British Telecom) in the adoption of switching systems in the 1960s and 1970s (B&M, pp. 196-198).

much else that rapid and widespread diffusion of the sort experienced with calculators and electronic watches is unlikely. No matter how great the benefits - and they are likely to be massive eventually - they would be exceeded by the costs of chaos occasioned by the immediate wholesale adoption of such revolutionary applications of semiconductor technology." (B&M, p. 95)

The Automobile Industry

Similarly to the case of the telecom sector, many observers saw the automobile as one of the greatest potential markets for semiconductor components already in the 1950s (B&M, p. 200). There is room for the use of such components in a large range of safety devices, in engine control systems (to reduce gas consumption and pollution, and increase engine efficiency), in monitoring, instrumentation and display, in servicing, etc. However, this potential was extremely slow to materialize.

A number of technical reasons made it difficult to incorporate semiconductors in cars: first, the conditions inside a car engine are very unfavorable to the introduction of microelectronic components, in that it is very hot, dirty and noisy. Second, components in a car must function for years, and it must be serviced by plain mechanics that had little knowledge of electronics. Third, while the prices of semiconductor components have declined steadily, that has not been the case for much of the sensor equipment needed to feed the information about the car's performance to the integrated circuits.

However, other sectors had faced even greater technical challenges and yet adopted semiconductors much earlier; indeed those difficulties fostered innovations that made adoption possible for a wider range of sectors. The real hindrance to early adoption in the automobile industry seems to lie elsewhere: given the massive capital sunk in existing manufacturing plants, any significant change in the design of cars, or in the type of components used requires extensive retooling, and changes to the production line that are difficult and expensive to implement. It is not so much what semiconductors could do to the cars themselves, but rather what they "threatened" to do to the production systems that became the decisive factor. Furthermore, the automobile industry in the US was at the time (the 1950s and 1960s) a tight, cozy oligopoly that resisted change and innovation in many dimensions, not just with respect to the new GPT. It was not until the late 1970s that the industry finally started to incorporate semiconductors in earnest, not only for mandated safety or pollution control features, and not just as optional features for upscale models, but for a wide range of monitoring and control functions in most models.[19]

6.3 Characterization of Early and Late Adopters

Recall that the key determinants of adoption in our model are, $n_{ic} = n_i^o/(G_i)^{\alpha/(1-\alpha)}$, and β_i/a_i. The smaller n_{ic}, the smaller the number of complementary components that have to be developed in order to adopt the new GPT, and hence the higher the likelihood of early adoption. The larger β_i/a_i, the larger the expected demand for products using the new GPT, relative to the development costs, making early adoption more likely. The discussion above suggests that the early adopters of the transistor can be characterized as follows:

(*i*) Very large G_is, due not so much to advanced capabilities of the new GPT per se (early transistors were quite rudimentary), but rather to the obvious deficiencies of the previous GPT in those particular uses.

(*ii*) In the case of computers, it was also a low n_i^o: electronic computers were born shortly before the transistor was invented, and hence there was not much time to perfect computers based on vacuum tubes.

(*iii*) Although we do not have direct evidence to that effect, it would seem that in both sectors a_i was low: since transistors were used in hearing aids and in computers

[19]In 1978 the estimated semiconductor component value of factory installations in automobiles stood at $111 million; by 1983 it had reached $620 million (B&M, p. 202). The semiconductor contents of cars have kept increasing rapidly since then.

primarily as substitutes for valves, incorporating them into the systems did not involve high development costs. It is when the introduction of semiconductors involve new functionalities, and/or the redesign of the systems in order to accommodate them that a_i's are high.

(iv) High willingness to pay for the relative advantages that the new GPT offered in those uses, by a (initially) small but well defined group of users. This translates into a (sufficiently) high and relatively certain stream of revenues that could at least cover production and development costs. Our model has just one demand parameter, β_i, and hence it cannot tell apart a high willingness to pay from say, a high level of demand. Thus, we just associate these favorable demand conditions with a high manifest β_i.

It is very hard to form solid judgments on the role that each of these factors may have played, but two issues are worth noting: the first is that it may not be a coincidence that for early adopters virtually *all* four parameters were particularly advantageous; in fact, casual comparisons with other sectors that adopted later suggest that it is indeed rare to find sectors for which this is the case. If pressed to single out one factor as particularly prominent, it would seem that it was (i), that is, the plain inadequacy of the incumbent technology vis a vis the new if still rough GPT. And then of course there is historical accident, which in this case takes the form of a remarkable coincidence for hearing aids: as a memorial to Alexander Graham Bell, who took a personal interest in the fate of the deaf, Bell Labs wavered royalties on the transistor patents to manufacturers of hearing aids. As to computers, the "accident" is in the timing, that is, the fact that they were developed just before the advent of the transistor, and hence benefited (in retrospect) from a very low n_i^o.

As to the laggards, it seems quite clear that the decisive impediment to early adoption was, in terms of our model, an extremely large n_i^o. In the case of telecommunications it was just the sheer size of the system itself and the complex interdependence of subsystems and components. In the case of automobiles the n_i^o referred to established manufacturing plants and the disruption that the introduction of new components could bring upon the production system. It is interesting to note that, in this context, past investments in the old GPT are not "sunk," since they determine the relative efficiency of the old versus the new GPT. The larger those past investments, the higher the efficiency threshold that lies in front of the new GPT in order to be adopted.

In retrospect, it is clear that the errors of forecast made initially with respect to the laggards were due to the fact that these predictions were made primarily on the basis of the expected G_is (in that respect they were right for both the telecom and the auto sectors), and not on the basis of n_i^o, or on the "composite" measure n_{ic}. By identifying these factors, our model helps at the very least think systematically about these complex and yet crucial set of issues.

References

[1] Braun, E. and S. MacDonald, *Revolution in Miniature - The History and Impact of Semiconductor Electronics.* 2d Edition (Cambridge: Cambridge University Press), 1982.

[2] Bresnahan, T. and M. Trajtenberg, "General Purpose Technologies — Engines of Growth?" *Journal of Econometrics* 65 (1), 1995, pp. 83-108.

[3] David, P.A., "Computer and Dynamo: The Modern Productivity Paradox in a Not-Too-Distant Mirror," in *Technology and Productivity: The Challenge for Economic Policy.* (Paris: OECD), 1991.

[4] Flamm, K., *Targeting the Computer.* (Washington, D.C.: The Brookings Institution), 1987.

[5] Griliches, Z., "Hybrid Corn: An Exploration in the Economics of Technological Change." *Econometrica* **25** (4), 1957.

[6] Helpman, E. and M. Trajtenberg, "A Time to Sow and a Time to Reap: Growth Based on General Purpose Technologies," NBER Working Paper No. 4854, 1994 (revised 1995).

[7] Judd, K.L., "On the Performance of Patents," *Econometrica* 53, 1985, pp. 567-586.

[8] Mansfield, E., *Industrial Research and Technological Innovation.* (New York: Norton), 1968.

[9] Williams, Michael R., *A History of Computing Technology.* (New York: Prentice Hall Inc.), 1985.

Figure 1: Sectoral Wave

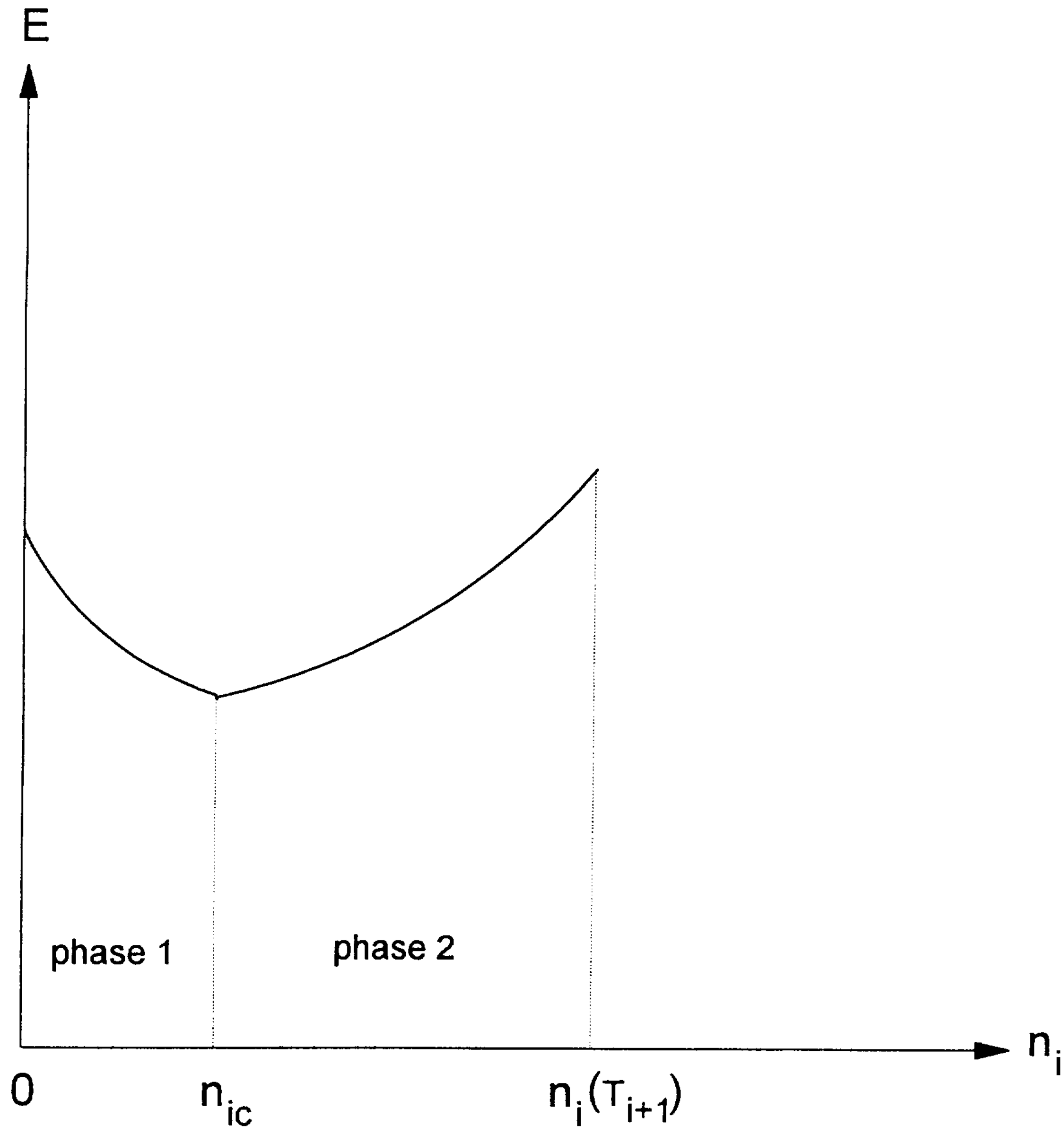


Figure 2: Steady State

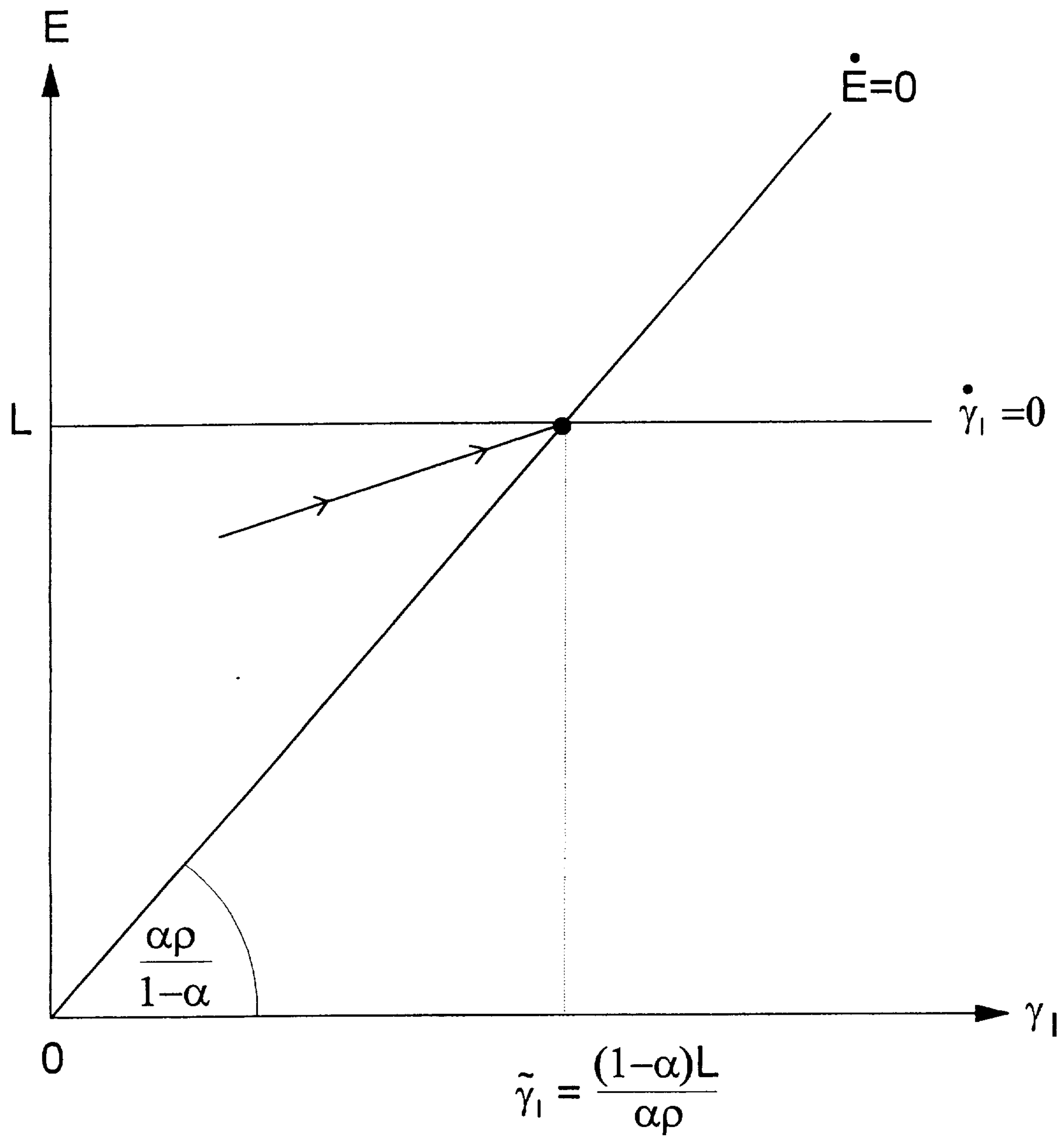
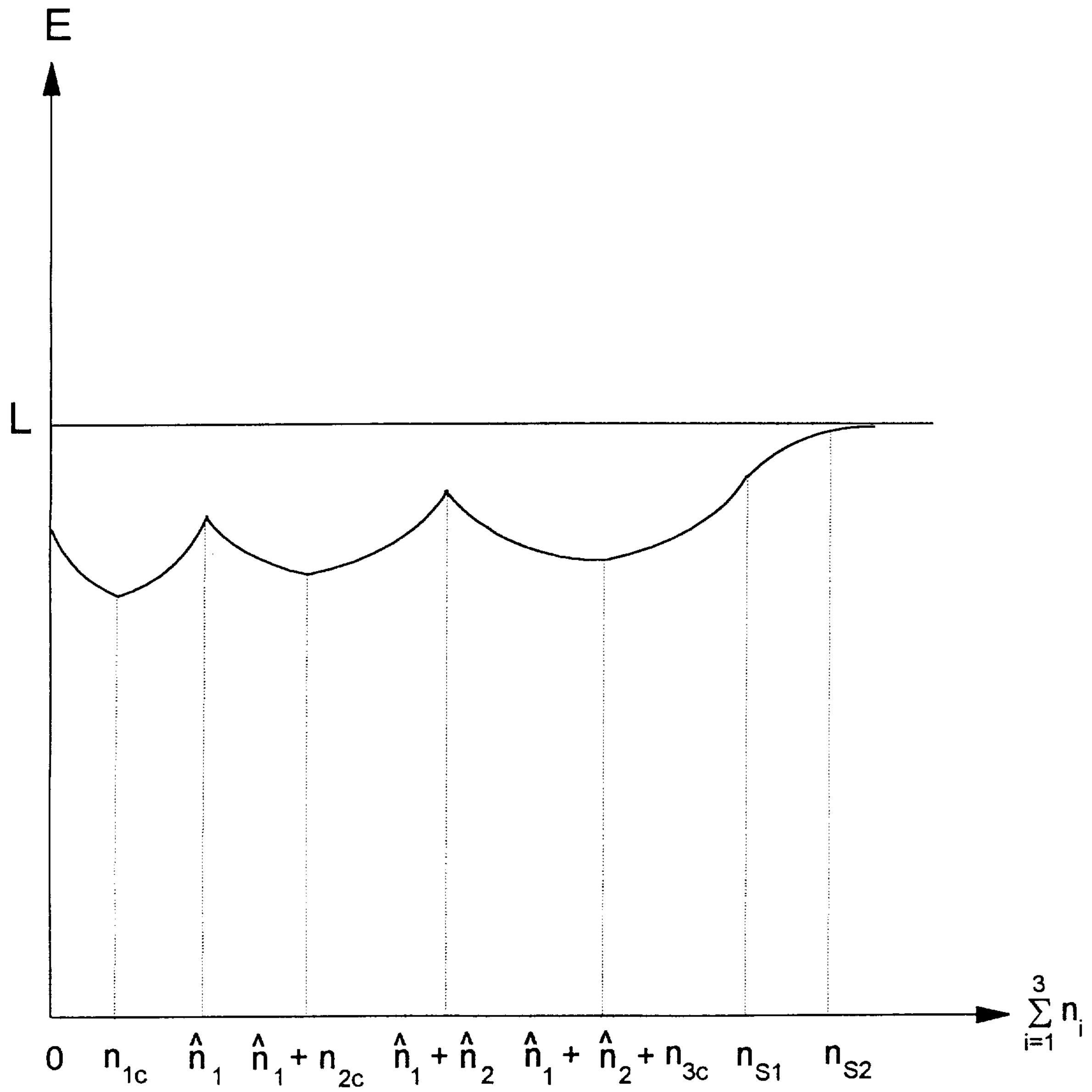


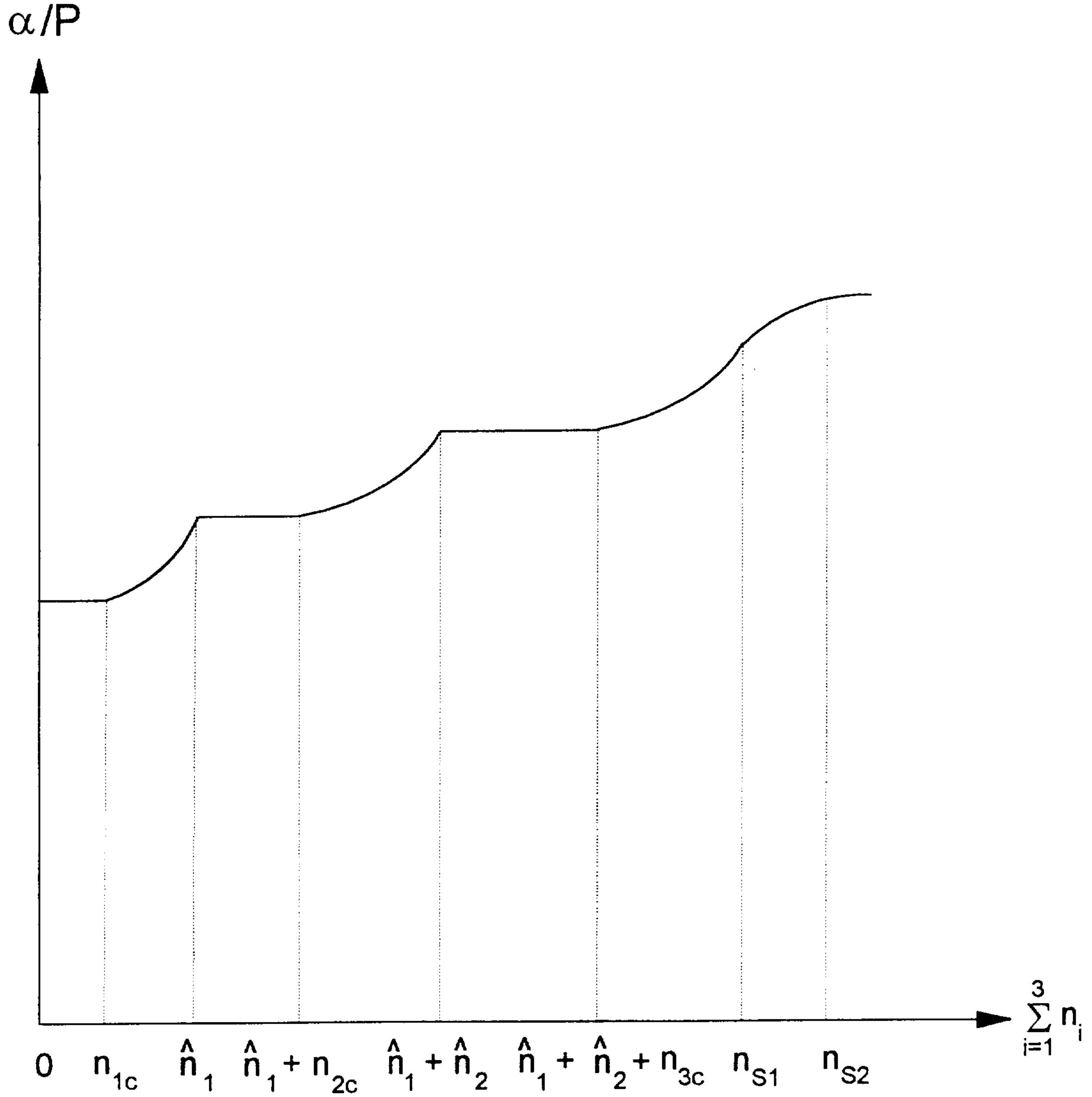
Figure 3: Equilibrium Trajectory



n_{S1} = one sector joins the second wave

n_{S2} = remaining sector joins the second wave

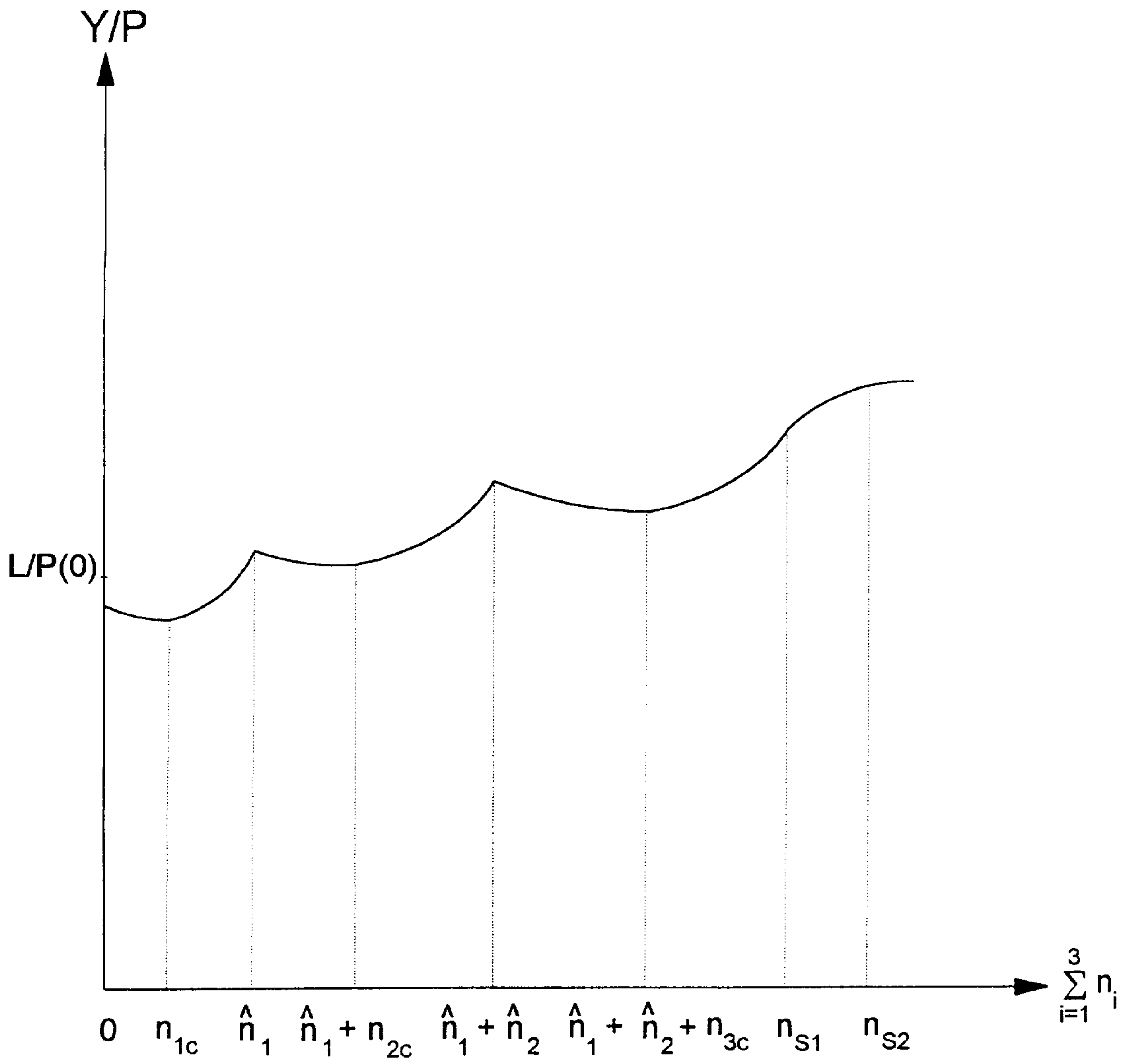
Figure 4: Real Wages



n_{S1} = one sector joins the second wave

n_{S2} = remaining sector joins the second wave

Figure 5: Real Income

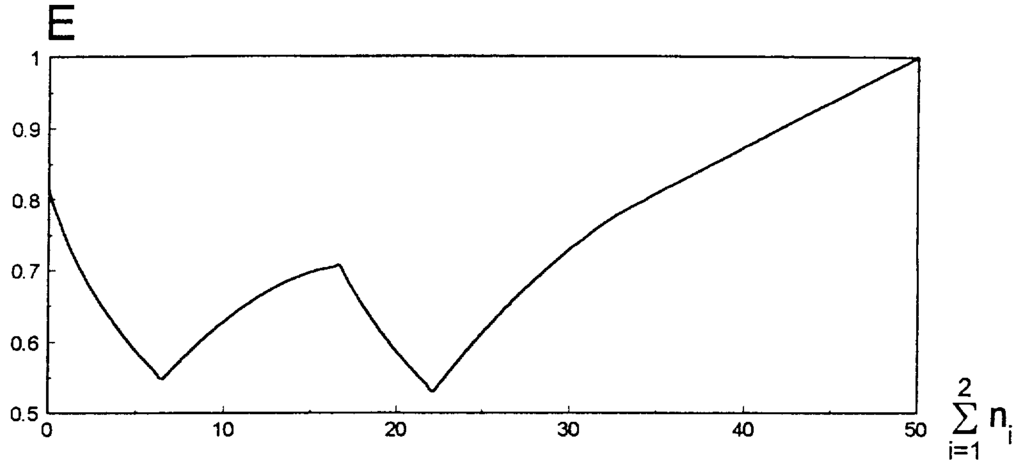


n_{S1} = one sector joins the second wave

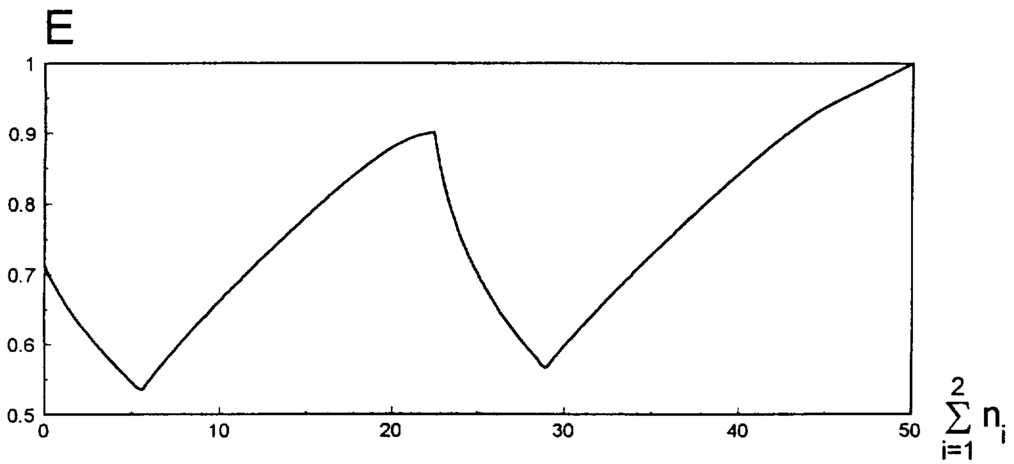
n_{S2} = remaining sector joins the second wave

Figure 6: Different Orders of Adoption

($L=1$; $a_i=1$; $\alpha=0.5$; $\beta_i=0.5$; $\rho=0.02$)



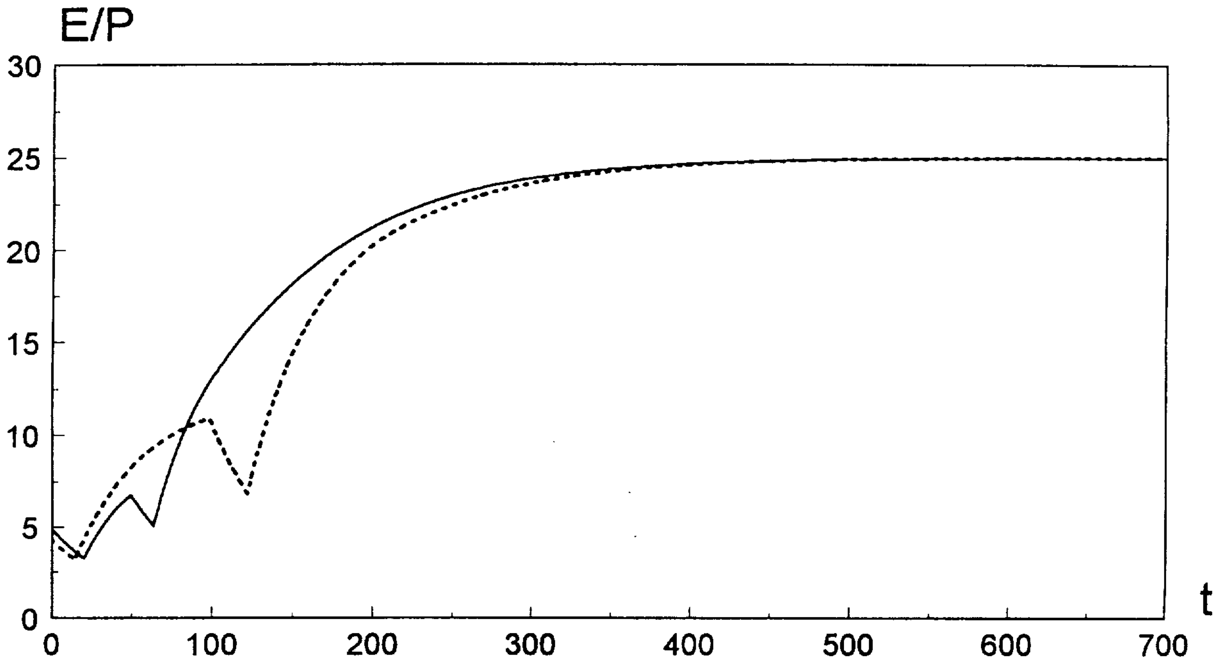
(a) $n_{1c}=6.5$; $n_{2c}=5.5$



(b) $n_{1c}=5.5$; $n_{2c}=6.5$

Figure 7: Welfare Comparison

($L=1$; $a_i=1$; $\alpha=0.5$; $\beta_i=0.5$; $\rho=0.02$)



solid line: $n_{1c}=6.5$; $n_{2c}=5.5$

broken line: $n_{1c}=5.5$; $n_{2c}=6.5$