

Technology Infrastructure in Information Technology Industries

W. Edward Steinmueller
Professor of the Economics of Technical Change
MERIT
Rijkuniversiteit Limburg
Maastricht, The Netherlands

March 20, 1995

Final Version

This paper has been prepared for Morris Teubal, Dominique Foray, Moshe Justman, and Ehud Zuscovitch (eds.), *Technological Infrastructure Policy (TIP): An International Perspective*, Kluwer Press, forthcoming 1995.

Both traditional and modern uses of the term infrastructure are related to "synergies," what economists call positive externalities, that are incompletely appropriated by the suppliers of goods and services within an economic system. The traditional idea of infrastructure was derived from the observation that the private gains from the construction and extension of transportation and communication networks, while very large, were also accompanied by additional large social gains. Thus, society as a whole had an interest in promoting these networks because they created new opportunities for economic choice and growth through "spillovers" that were involuntary in the sense that they could neither be avoided nor entirely captured by the creators of transport and communication networks.¹ Over the past century, publicly regulated and promoted investments in these types of infrastructure have been so large, and the resulting spread of competing transportation and communications modalities have become so pervasive, that they have come to be taken as a defining characteristic of industrialized nations. The size and pervasiveness of these traditional infrastructures is now so great that they no longer command the social attention and priority that previously supported their public subsidization and regulation.

During the past half century, a new type of infrastructure has come to be recognized as important within industrialized nations.² This new infrastructure is based on knowledge that spans the boundaries of individual enterprises and that originates either from explicit efforts to create spillovers through the public disclosure of knowledge or, implicitly in involuntary knowledge spillovers from the commercial activities of enterprises. As with earlier infrastructures, the infrastructure arising from knowledge creates private as well as social gains. Using knowledge requires "capabilities," business

¹ The "involuntary" character of spillover benefits has, of course, also been the result of public regulation of the monopoly profits of transport and telecommunication industries. See McGraw [1984] for an outstanding history of the historical development of such regulation.

² The watershed contribution in justifying the extension of public support for basic science was "Science: The Endless Frontier," Bush (1945). More recently, Tassef (1992) has articulated a vision of how technology infrastructures are characteristic of modern "system" technologies. Both Bush and Tassef are concerned with distinguishing clear boundaries between public and private roles.

enterprise skills and experience in adapting and absorbing knowledge, that are used for the development of new or the improvement of existing products and production processes.³

The form of knowledge is also relevant to this discussion. Knowledge may be codified as in the case of operating procedures, patents, or blueprints, or it may remain more tacit and intrinsic as in problem solving experience or accumulated skill that distinguish the real world enterprise from the abstract firm most often encountered in economic analysis.⁴ Although both codified and tacit knowledge has the potential for creating positive spillovers across enterprises, the mechanisms by which these different types of knowledge are disseminated and received differ. Because codified knowledge may be widely and quickly disseminated, property right mechanisms are necessary for preserving the creator's economic interest in such knowledge. These property rights form the basis of market transactions in the exchange of such knowledge and preserve the incentives for its creation and dissemination. Tacit knowledge, because it is more difficult to disseminate, may be more readily linked to other forms of business relationship such as joint ventures and technology exchange agreements. Both types of knowledge, however, signal competitors that commercialization opportunities exist and provide knowledge that may be useful for developing substitutes for available knowledge. Moreover, when the technological foundations of an industry facilitate a variety of technological solutions, the advance of knowledge is likely to be broadened and diversified through processes of adaptation, imitation, and reproduction outside the boundaries determined by intellectual property rights or contractual relationships. The major theme of this chapter is that advances in information technology have created knowledge spillovers that have opened more opportunities for competitive entry than they have closed.

³ There has been considerable discussion in recent years of the idea of "core" competences, e.g. Teece (1987) and Prahalad and Hamel (1990), which are capabilities that are difficult to transfer or reproduce and that therefore serve as stable anchoring positions for investment in business enterprise.

⁴ Winter (1987) offers a very useful introduction to the distinctions between and consequences of tacit and codified knowledge.

Infrastructure

From the viewpoint of a society, knowledge creates positive externalities within national economies or, under conditions of liberal global trade and scientific exchange, within the worldwide economy.⁵ For our purposes, there are two key features of knowledge that make it part of the technological infrastructure of an industry. First, the externalities it creates are incompletely captured or appropriated by individual enterprises, or even larger networks of firms, so that it may be said that they are one dimension of an economies' social capabilities.⁶ When knowledge is incompletely appropriated it is possible to distinguish between private and social gains. When knowledge is effectively owned and only transferred through market transactions, there is little to distinguish it from other specialized assets that, within a competitive framework, should earn returns commensurate with their marginal revenue product.⁷ Our concern, however, is with knowledge that makes a contribution to social capabilities above the returns earned by its creators or owners.

⁵ Many of the proponents of capabilities or "core" competences are particularly concerned about the consequences of erosion of these capabilities within particular domestic economies, a signal of which is often purported to be the use of foreign produced inputs (e.g. Cohen and Zysman [1987]). It is nonetheless true that the pattern of international trade among advanced industrial economies is increasingly based upon intra-sectoral exchange, a pattern that could only exist if there was intra-sectoral specialization among trading partners, i.e. the development of complementary capabilities for similar economic activities. Concluding that this pattern of trade, or even the current liberality of world trade, will continue indefinitely may, however, still provide a basis for concern.

⁶ The term "social capabilities" has been employed by Abramovitz [1989] in explaining national differences in long term economic growth performance. His much broader view of social capabilities encompasses institutional, cultural, and educational factors that contribute a nation's ability to catch up or overtake other nations in economic growth. Here, we will consider only the problems of gaining access to technology within the broader issue of social capabilities.

⁷ Nonetheless, the commercial availability of specialized inputs is an important dimension of social capabilities, even if the returns from these inputs are privately captured.

The second feature of infrastructural knowledge is that it persists over time. While knowledge, in general, may be constructed and dismantled rapidly, knowledge that plays an infrastructural role in industries needs to persist long enough that it can be recognized and exploited by enterprises and other organizations not directly involved in its creation. In other words, confining attention to knowledge that is persistent is a way of focussing attention on attributes or characteristics of knowledge that take time for individuals and organizations to acquire. Moreover, investments in knowledge acquisition provide an incentive to find new applications for this knowledge after it has been acquired. Persistence is thus linked to the cumulative and incremental character of technological progress.

In the information technology industry, several types of knowledge, play an important infrastructural role in being incompletely appropriable and persistent. Information technology industries, the collection of enterprises offering products and services based on electronic methods of acquiring, storing, processing, and distributing information, have benefitted from knowledge advances in integrated circuit, optoelectronic, and magnetic technologies. In addition, each of these "enabling" technologies has important roots in materials science.⁸ Both the enabling technologies and materials science have the features of incomplete appropriability and persistence that are necessary for the emergence of a technological infrastructure. Thus, one approach to the problem of

⁸ In Tasse's framework, materials science, including solid state physics, is the most "generic" technology underlying information technology, see Tasse [1992], pp. 98-99. Tasse also uses the term "infratechnology." Tasse defines infratechnologies as "practices and techniques, basic data, measurement methods, and measurement-related concepts which increase the productivity or efficiency of each phase of the R&D, production, and the market development stages of economic activity," Tasse [1992], p. 99. Tasse asserts that infratechnologies are "generally competitively neutral, in large part because of their typical widespread and uniform use." If, however, such technologies are a sufficient technological basis for entry, they cannot be competitively "neutral" but will enhance competition. Unfortunately, within Tasse's framework, technologies are either "proprietary" (i.e. subject to ownership) or "infrastructural" (i.e. "generally" competitively neutral and widely available). Involuntary spillovers of knowledge that enable entry are both available and competitively non-neutral (i.e. they are competition enhancing). Thus, while such spillovers fit Tasse's epistemological definition, but they fail to conform to his behavioral hypothesis (competitive neutrality).

information technology infrastructure is to examine how these technologies have developed.⁹ The common technological foundations of information technologies have contributed to a blurring of boundaries among different types of information and thus between different segments of the industry. Streams of data once associated with large and remote computers have become linked through common information storage, processing, and transmitting technologies bringing together data streams representing voice, still and moving images, and text. Increasingly, all of these data streams flow over a global network of telecommunications links and through a forest of computers, large and small, near and remote. This process of convergence in the *use* of the technology has created an important knowledge pool that is distinct, yet related, to the knowledge used in the production of information technology products.

The significance of technological convergence suggests two complementary approaches to the analysis of technological infrastructure. One approach is to pursue the origins of the knowledge of materials and enabling technologies supporting technological convergence among the information technology industries.¹⁰ The other approach focusses on the creation of common pools of knowledge in the use of these technologies. This chapter focusses almost exclusively on the latter approach since it is the use of new technology where the experience of the information technology industries can offer a unique contribution to our understanding technological infrastructure. Moreover, it is the use of these technologies that provides many of the inducement mechanisms for generating innovations. This choice of focus leads, however, to an important qualification. The enormous, perhaps unprecedented, technological advances in materials and facilitating technologies, combined with their wide breadth of applicability, are an important part of

⁹ Moreover, many industry observers would concur that the boundaries of the information technology industry should be further enlarged to include enabling technologies produced by segments of the scientific instrument and capital goods industries that permit the fabrication, testing, and quality control of these materials as well as complementary technologies such as electronic displays and fibre optics that are also closely linked to new materials.

¹⁰ Pursuing this first approach would likely lead to a greater emphasis on what will later be termed the "traditional" approach to understanding knowledge-based infrastructure in which publicly-funded scientific advance plays a central role.

the explanation of the evolution of knowledge demand and utilization in the information technology industry. Caution is required in applying lessons from the experience of the information technology industries to industries where technologies are more heterogenous and subject to slower and less uniform rates of advance.

The idea that knowledge is infrastructural is largely a contemporary construct reflecting the growing importance and complexity of knowledge used in the production of goods and services. Earlier ideas, such as the "industrial district" of Alfred Marshall or even the division and specialization of labor depicted by Adam Smith, captured important elements of knowledge as infrastructure. Two modern trends in industrialized economies differentiate the modern use of the term technological infrastructure from earlier externality-generating developments such as industrial districts. The first trend is the growing internationalization of the "public good" character of knowledge, i.e. its availability throughout the world to an ever larger number of organizations.¹¹ It is increasingly likely that research developments, wherever they are achieved, will be monitored and imitated somewhere else in the world. This trend applies not only to research knowledge created as the result of public funding, where the social value of knowledge is now well-established, but also to knowledge developed in the private sector. Commercially generated knowledge creates voluntary and involuntary spillovers that also have public good features.¹² These spillovers from commercial research activities are available globally, subject to the problems of adaptive and absorptive capacities as well as intellectual property restrictions.

¹¹ The other characteristic of public goods, non-rivalrous consumption, i.e. the possibility, within limits, of increasing consumption without affecting the abilities of others to consume, is an inherent characteristic of knowledge. Knowledge is a "purer" public good in this respect than textbook examples like clean air or water that are subject to congestion externalities.

¹² The economic analysis of involuntary spillovers has received considerable attention recently. See van Meijl [1994] for a review of several contributions as well as a model of the impact of such spillovers on the R&D investment choices of firms. Rosenberg [1990] offers a number of reasons that firms perform basic research with their own funds. Several these reasons, including participation of company researchers in scientific communities that require public disclosure of research results, encourage "voluntary" knowledge spillovers.

The second trend is the increasing importance of capabilities for imitating, adapting, reproducing or independently recreating knowledge first "discovered" or "generated" elsewhere. This trend is directly related to the growing commercial use of scientific and technical knowledge as an input in product and production process innovation.¹³ Because a vast amount of knowledge has become available, the relative importance of capabilities for utilizing and accessing this knowledge has increased so much that it may now be said that capabilities to access knowledge are essential to its infrastructural character.¹⁴

The growing significance of capabilities for accessing knowledge suggests a reorientation of the analysis of technical change toward the study of barriers and enabling mechanisms for such access.¹⁵ The same re-orientation is implied for the study of infrastructure. Thus, despite the continue importance of infrastructural investments in generating new knowledge as a public good, more attention should be devoted to barriers and enabling mechanisms for knowledge utilization. This chapter examines how technological trajectories and technical interface methods contribute to the creation of a technological infrastructure in the information technology industry. Other mechanisms, however, in the areas of intellectual property and methods for financing technological progress deserve equal standing in a complete picture of the technological infrastructure of the information technology industries.¹⁶

The infrastructure created by the public good character of knowledge has an influence on industrial structure. In particular, the extent of asymmetries in the

¹³ Kuznets [1966] argued that the development of "science-based industry" was an epochal change, distinguishing our modern era from the past and unrivaled since the development agricultural cultivation in its impact of economic growth.

¹⁴ See Cohen and Levinthal [1990] and Barabaschi [1992].

¹⁵ See David and Foray [1994].

¹⁶ For example, see Mowery and Steinmueller [1994] on the effect of the 1955 consent decree which led to the release of AT&T's intellectual property in semiconductor technology. The ability to finance new enterprises has a major impact on whether existing enterprises can extend their control over new knowledge without challenge from competitive entrants.

accumulation of knowledge and capabilities seems to be one determinant of patterns of competitive entry and the ability of incumbents to retain a technological lead over new entrants. In several segments of the information technology industry--computer sub-systems, personal computers, and semiconductors--substantial competitive impetus comes from new entrants. Competitive entry highlights the nature of "involuntary" spillovers from commercial activities that create a technology infrastructure without explicit coordination mechanisms. Although accumulation of capabilities through involuntary spillovers, entry, and lack of explicit coordination is one means of stimulating technological progress and competitive fitness, it is not necessarily the only method. Other industry segments such as telecommunication switches, large computer systems, and massive software projects may require knowledge accumulation *within* a single large enterprise. Moreover, it is clear from international comparisons that different industry structures may generate comparable innovative capabilities.¹⁷ As a result of these qualifications, this chapter does not attempt to create a normative theory of what industry structure is best suited for exploiting the infrastructure that arises from involuntary spillovers of knowledge. Instead, it has the more limited purposes of demonstrating how such infrastructures come to exist and tracing some of their impacts.

To summarize this introductory discussion, the traditional idea of infrastructure has proven enormously useful in mobilizing public support for constructing transportation and communications links and, more recently, for the public funding of research within industrialized nations. While infrastructures are always characterized by positive externalities, they are not necessarily entirely based upon publicly funded research that leads to public goods. Knowledge that is developed in the private sector is often imperfectly "appropriated" by its creators and spills over to other enterprises and these spillovers persist over a long enough time so that they allow slower moving processes like the creation of new enterprises to operate.

¹⁷ See Steinmueller [1988] for an argument that this is the case with the U.S. and Japanese semiconductor industries.

The main theme of this chapter is that technological infrastructure in several of the information technology industries involves substantial contributions from the involuntary spillovers of knowledge in the process of creating and using such technologies. This method is illustrated with examples drawn from the computer and integrated circuit industries where significant involuntary spillovers of knowledge have influenced technological progress and industry structure. The implications of these examples for our understanding of the influence of infrastructure on horizontal market structure and competition policy are explored from a general viewpoint. Vertical market structure influences of infrastructure are also examined with a further example highlighting the role of interface compatibility standards. The chapter concludes by identifying some unresolved issues and qualifications that suggest new directions for research and policy.

A Retrospective View of the Development of Infrastructure in the Computer Industry

Whatever page one selects from the early history of digital computers, one finds a solitary researcher or a team working in a university setting and with government support to translate the abstract idea of a stored program digital computer into working electronic circuitry.¹⁸ This history is, however, subject to some important qualifications. While one may trace the path of development of the digital computer from the university research laboratory to the commercial products of Burroughs and IBM, the latter company had, during the 1920s, developed commercial applications for automated data processing using the electro-mechanical technology of tabulating card punches, readers, sorters, collators, and accounting engines. Among the legacies of this earlier history was the use of tabulating cards as input and output media. For IBM, however, the earlier experience had provided a more important advantage--a unique insight into how data processing equipment could be utilized in the business environment. The integration of previous data-processing methods with the use of digital computers highlights an under-appreciated

¹⁸ See Flamm [1987] and [1988] for an economic history and Ritchie [1986] and Williams [1985] for a more general histories of these developments.

characteristic of technological change is that the initial implementation of new technologies have little relation to their ultimate use.¹⁹

The computer, as developed with government support, was essentially a scientific instrument serving as a substitute for human "computers" in the calculation of reference tables for scientific and military application. Transforming this scientific instrument into an engine for the everyday processing of data streams originating from the operation of business enterprise is now seen as an obvious and "derivative" application of computer technology.²⁰ In practice, however, this application required a decade of experimentation before it could be systematized in the development of the IBM/360 and the competing mainframe computer offerings of IBM's rivals.²¹

IBM had learned from its previous experience that commercial applications of dataprocessing were largely focussed on "data reduction" rather than "data generation." Much of the use of computers as scientific instruments was devoted to the solution of discrete approximations of continuous mathematical functions that generated an output stream of data larger than the input stream. By contrast, in commercial applications, the incoming stream of data is larger than the output stream. The punched card, which had played a secondary role to paper tapes or direct "hard wiring" techniques for programming early computers became the basis for IBM's commercial computers because it was well

¹⁹ Rosenberg [1976] makes the general point while Ceruzzi [1986] develops the specific case of the computer industry.

²⁰ See Ceruzzi [1986]. Ceruzzi quotes one of the pioneers of the computer industry, Howard Aiken, as claiming "...if it should ever turn out that the basic logics of a machine designed for the numerical solution of differential equations coincide with the logics of a machine intended to make bills for a department store, I would regard this as the most amazing coincidence that I have ever encountered," Aiken [1956].

²¹ See Pugh, Johnson, and Palmer [1991].

suited (and previously developed) for the creation of large scale input data streams and thus had a persistent influence in the development of commercial digital computers.²²

In addition, the data reduction role of computers suggested a particular organizational model for their use. A commercial enterprise often generates data through a geographically extensive network of operations. Data processing activities required centralization to properly utilize the capacity of the relatively expensive digital computer, but centralization was also the consequence of past experience in the commercial operation of data reduction in data processing applications in banking, insurance, and other commercial applications of data processing. In these applications, media (such as punch cards magnetic tape, or optical readable paper forms) were physically transported to central locations for data processing. This operating procedure generated important "bottlenecks" whose expansion would offer technological improvements as data processing was further automated using the digital computer. One method of widening the bottleneck of physical transport of paper input streams was to decentralize the input of data through the use of "remote job entry" stations connected through telecommunication links. In other words, the bottlenecks of centralized data collection provided an incentive for the development of new computer hardware that would reduce costs, speed transfer, and increase the reliability of the data stream. The existence of a bottleneck served to define a standard of performance for existing technologies by which alternatives could be assessed.²³

The central role of the punch card as input medium and "data reduction" as operating procedure have important features of an infrastructure. The punch card provided a known standard for the encoding of streams of data while the problems of centralized data collection and data reduction provided a comprehensible target for innovative efforts

²² For three decades, most of the data that flowed into large computers was packaged in columns of 80 characters, the length of a punched card, and even today many data display terminals are designed with displays 80 characters in width.

²³ See Rosenberg [1976] for a discussion of the relationships between technological bottlenecks and trajectories of technological improvement.

as well as a variety of defined interfaces or links among sub-systems. Moreover, the incorporation of these media into organizational routines meant that their use was persistent and growing over time as more commercial enterprises adopted these methods.

The variety of technological solutions created within this framework, in turn, created a foundation for further technological improvements. "Remote job entry" stations would become interactive data terminals where data could be entered directly under the control of "time shared" central computers and the physical transport of data entry media was often replaced by increasingly sophisticated telecommunications links for the transfer of data. These improvements were, to an important extent, motivated by the bottlenecks of data transfer and served to define a trajectory for improvements. They were also only one of the bottleneck and improvement trajectory mechanisms that operated in the commercial development of digital computers. Others that were important were the development of minicomputers for real time control of industrial equipment which widened the bottlenecks of input channels to the large computer, the development of faster and higher capacity mass storage devices such as the magnetic disk which served as a means of widening the bottleneck of a fast and reliable temporary storage method for program instructions and data, and the continued improvement of magnetic tape storage devices which widened the bottleneck of the punched card itself by offering inexpensive and faster mass storage. Each of these subsystems were connected to digital computers through interface specifications defined by computer manufacturers but that nonetheless became points of entry for competitive products.

Integrated Circuits and Digitization

The invention of the integrated circuit (IC) in 1961 opened another important trajectory in the information technology industries. The IC was initially a "niche" device whose early applications, such as the guidance control units for intercontinental ballistic missiles, justified a relatively high cost per electronic logic function to deliver reliability and miniaturization performance. During the 1960s, this cost fell dramatically, so that by the end of the decade, a new company, Intel, was formed with the aim of developing a solid state computer memory that would be price competitive with magnetic core memory

sub-systems for mainframe computers. Like the data reduction trajectory described above, the trajectory of IC improvement served to focus attention on the opportunities for new applications. The bottleneck in this case was the cost and time required to develop electronic systems.

One of the most influential developments in conversion of electronic signals to streams of binary data or "digitization." While digitization was already a well-established principle in the design of electronic systems, it gained enormous impetus from the cost reduction of logic elements in integrated circuits. A leading authority of the mid-1970s summarized this development by the maxim that it "invalidates the traditional goals of logic design," most of which were related to the minimizing the number of circuit elements.²⁴ The number of such elements had to be minimized because the costs of the electronic system were directly proportional to the number of electronic components required to perform a single operation.²⁵ With transistor technology a handful of components were required for every elementary application; with the first ICs dozens of operations could be performed with a single component, and soon thereafter thousands of such operations could be performed by a single IC.

The particular set of skills associated with digitization have become a central feature of electronic system design across a very broad spectrum of applications. Where electronic system designers previously approached each design problem as having unique requirements that required an integrated design, digitization has allowed the designer to compartmentalize design problems into stages or sub-systems that communicate with one another using a relatively limited set of interface methods. The set of design skills, and the menu of available IC devices for implementing designs, has accumulated steadily for the past two decades so that the designs of almost all electronic systems, regardless of

²⁴ Blakeslee [1975], p. 15.

²⁵ See Braun and Macdonald [1982], p. 99 for an excellent discussion of this problem whose severity set definite technological limits to transistor-based electronic systems.

level of complexity, are now implemented using such design methods.²⁶ These techniques are likely to persist indefinitely due to the convergence of information transmission and storage methods noted in the introduction to this chapter. Digitization has had a marked effect in creating a common focus for the product and process innovations of the integrated circuit industry, and has been particularly important in achieving user acceptance of new products. The technical compatibility of new products is greatly facilitated by the evolution of a common approach, digitization, to the construction of electronic systems. This approach has created both substitution and complementarity relationships among ICs regardless of their manufacturer.

Digitization has thus created the sort of knowledge spillover from commercial activities that we have identified as infrastructural in being both persistent and creating widespread involuntary knowledge spillovers. A new entrant may examine the range of available products and the possibilities for new applications and devise a product innovation strategy that fills a new application need or establishes a new market niche, often without any need to be concerned about claims of intellectual property since the variety of feasible techniques is very large.²⁷ Despite the strong position of existing producers, the process of entry in the IC industry has continued, at a varying pace, for over two decades and despite recurrent predictions that additional new entry will be foreclosed by entry barriers from accumulated experience, the growing costs of state of the art facilities, or the profusion of available product solutions.

²⁶ This development was summarized by one industry pundit, Nicholas Negroponte, with the aphorism; "There will be many more MIPS in the nation's appliances than in its computers." (MIPS is million-instructions-per-second, a standard measure of computer power). Brand [1983], p.5.

²⁷ The technologies for manufacture of ICs are, however, better protected and entry increasingly requires a strategy of innovation to gain access to established knowledge, see Mowery and Steinmueller [1994].

Implications for Our Understanding of Infrastructure

The examples from the technological history of computers and integrated circuits have important implications for how we view the economics of infrastructure. After its initiating role, public investments in the development of digital computer technology were displaced by evolution along a "trajectory" of improvement defined by the "focussing devices" of data reduction underlying the commercial application of digital computers and the interactions among the subsystems comprising the digital computer. Throughout these developments, public funding continued to play important roles in the creation and standardization of programming languages, the development of data communications technology, and the improvement of digital computers through the use of transistors and integrated circuits. Nonetheless, the rate and direction of innovation in the computer industry was powerfully influenced by the logic of "data reduction" and subsystem integration which were fundamental characteristics of commercial data processing applications and the evolution of the digital computer as a "system" with accompanying bottlenecks and complementarities.

Similarly, the early funding of IC research from the sales of high performance devices to the U.S. Department of Defense set in motion the development of the IC as a practical commercial device and provided the U.S. with important first mover advantages in this technology. Nonetheless, the subsequent history of the IC industry has demonstrated that this first mover advantage was insufficient to preclude entry, either by strong foreign competition or by a succession of domestic firms in the U.S.²⁸ The inability to retain a first mover advantage was the outcome of several developments, two of which were the development of a limited number of interface methods for linking ICs together to make electronic systems and the, related, growing significance of digitization as a design principle in the creation of electronic systems. Digitization has permitted users (electronic system designers) to rapidly understand and accept new product innovations and has provided a framework for product innovation.

²⁸ See Steinmueller [1988], Okimoto, Sugano, and Weinstein [1984] and Braun and Macdonald [1982] for accounts of this history.

From an economic perspective, the knowledge generated from attempts to manage data streams in order to achieve data reduction and to advance digitization as a principle of electronic system design were very imperfectly appropriable. The requirements for standardizing data flows and creating "compatible" subsystems were the points of entry for competing products in both the computer and IC industries. Knowledge of the performance of the existing techniques was available to anyone that had experience with actual commercial applications. In short, the broad diffusion of technology brings with it a widespread knowledge of the limits of existing solutions and the corresponding opportunities for improvement. It is therefore not surprising that this process should encourage that the entry of enterprises offering improvements within the framework of existing methods.

The inability of incumbents to dominate these new developments is largely explained by the dynamics of product improvement. New innovations often appear in implementations that are only tangentially related to the eventual implementations based upon incremental improvement. For larger enterprises, the landscape of technological change is often defined by opportunities for significant increments to their sales and profits. In practice, however, few innovative products are immediately able to deliver very significant increments to revenues or profits. Innovations are therefore continuously born at the periphery of economic significance where the largest firms are unlikely to employ their advantages of scale and experience. Experience with these innovations, in turn, create localized knowledge that becomes a source of first mover advantage when such products become successful in the market. This mechanism reverses the competitive advantage of incumbent and entrant, significantly increasing the uncertainty that the incumbent will catch up with the entrant. The risks of the entrant, though still large, are substantially reduced by the existence of accepted solutions and defined trajectories of improvement of these solutions. In the computer industry, these possibilities for entry were enormously facilitated by the modularity of subsystems and the definition of interface methods by which these subsystems could be integrated into larger systems.

Interface methods for achieving technical compatibility are enormously important in the information technology industries as a means of flexibly combining diverse

collections of components and sub-systems into information processing systems. Interface methods create spillovers that are imperfectly appropriated by their designers and that become entry opportunities for other firms. The entrant benefits from the existence of interface methods since they define a point of market entry and a target for innovative efforts. The range of entrant's innovative opportunities is, however, likely to be constrained by the fact that specific technological solutions, including interface methods, are accompanied by particularities that are implicit or tacit and that only emerge as barriers when entrants attempt variations. For example, the creator of an interface method may or may not choose to anticipate possible improvements in the performance of devices that may be connected to such an interface. One approach to regularizing interface methods is to develop a voluntary standard, a publicly available specification of the interface method, that attempts to completely define the technical requirements necessary for compatibility. In the framework of this chapter, the creation of a voluntary standard for interfaces is a case of transforming an involuntary to a voluntary knowledge spillover. The desirability of this conversion for the interface designer depends upon whether such a voluntary standard will increase their revenues in products embodying or purchased as complements to the standard.²⁹

A voluntary standard may not be chosen in an effort to profit directly from the control of the interface method. The "strategic" use of interface methods has a mixed history. During the 1970s, IBM was accused of manipulating interface methods to disadvantage competitors without major harm to its own sales of computers. During the 1980s several companies including Texas Instruments sought control of the technical interface between software and personal computers with disastrous consequences to their computer products. More recently, computer game manufacturers such as Nintendo and Sega have sought to control interface methods for their products without apparent harm to their commercial success. It is difficult to conclude from this experience what choice a business enterprise should make with regard to control of interfaces. Similarly, the public welfare implications are uncertain. Control of compatibility interfaces is a means of appropriating quasi-rents that might otherwise be competed away by competitive

²⁹ See David and Steinmueller [1994] for further discussion.

opportunities, but these quasi-rents may finance innovative improvements that deliver higher levels of public welfare in the future.³⁰

Implications of Technological Infrastructure for Industry Structure

The externalities created by involuntary knowledge spillovers that are a central feature of important segments of the information technology industry have implications for industry structure beyond the specific issues of interface methods. With respect to horizontal structure, involuntary spillovers must be examined in relation to the economies of scale that are typical in the information technology industries. In examining vertical structure, the existence of spillovers complicates the competitive process by creating the potential for vertically related firms to vie for profits from the control of markets or technological leadership. The existence of a common infrastructure reduces the determinacy of vertical structure, bringing "component" suppliers into competition with "system" producers because the effect of the common infrastructure is to blur distinctions between components and systems. The horizontal implications of the technological infrastructure are examined first, and in general terms. This is followed by a discussion of vertical implications using the example of personal computers.

Cost reduction through accumulated production experience, or learning, is a common feature in manufacturing industries generally, and many of the information technology industries in specific. While learning often provides advantages to incumbents that can exploit their previous investments in building sales and distribution networks to

³⁰ The term quasi-rent is used here to designate the existence of economic profits in excess of competitive returns that exist by virtue of an "artificial" barrier to competitive entry. In the case of interface compatibility standards it is assumed that market incumbents may disadvantage potential entrants without doing themselves serious damage in competition with other incumbent producers, i.e. their choice does not lead to a marked deterioration of the technical quality of product available to purchasers. In addition, it may be the case that an incumbent has accumulated market power relative to rivals through mechanisms other than the technical qualities of its product such as superior sales and marketing efforts. In this case, the use of interface stations to generate quasi-rents is an extension of the market power originating from other sources.

rapidly increase cumulative output, learning will also offer cost advantages to any firm that is able to rapidly increase its share of a market relative to competitors. Static economies of scale occur when increases in current output result in less than proportionate increases in total cost so that the average cost of each unit of output is declining. In information technology industries, both dynamic and static economies of scale are often present. Learning related advantages arise from the reduction of the costs of making continued progress through the incremental process described above as well as the accumulation of knowledge of how components can be refined and systems more efficiently integrated. Static scale economies are most often the direct result of spreading fixed product development costs over larger output.³¹

"Unexploited" economies of scale may suggest a public welfare motive for limiting the extent of competition. After all, a single firm can achieve higher levels of cumulative or current output and therefore lower costs, so why not limit competition to assure the greatest possible cost reduction? There are two answers to this question. The first, and simplest, applies to the case of those parts of the information technology industries where product differentiation plays a modest role in competition. Under these conditions, there is a direct tradeoff between the advantages of learning or other dynamic scale economies and the creation of monopoly power and pricing. The impact of "involuntary" spillovers is to limit the realization of such scale economies and it may therefore seem to be in the public interest to limit the extent such spillovers. On the other hand, if products are differentiated, as is most common in the information technology industries, it is unclear which product should receive the benefits of scale--choosing one may lead to an inferior outcome for consumers or disadvantage the industry in international competition. This suggests a negative answer for the question; limiting competition will reduce the variety of products and is therefore likely to provide a solution that is sub-optimal from a social

³¹ While the spreading of fixed costs is, properly speaking, a static economy of scale, it is often only realized through links to dynamic processes such as positive feedbacks that occur through the adoption process. One common feedback mechanism is that although current period demand is limited, demand over a longer term shifts outward as many users adopt the product and the externalities from the accumulation of specific skills and complementary products encourage further adoption.

welfare viewpoint. The horizontal market structure implications of involuntary knowledge spillovers therefore depends upon the value of product variety--the greater the value associated with variety in terms of customer demand or competitive fitness, the more likely it will be that horizontal structure will be unconcentrated.

Involuntary knowledge spillovers have had a major impact on the vertical structure of the personal computer industry which has experienced a shifting vertical locus of innovative activity. IBM initially sought to establish a commonly recognized standard for the use of microprocessors in small computer systems by adopting the Intel 8088 microprocessor which was also "second sourced" from AMD and eventually other companies. The innovative lead passed upstream to Intel and Microsoft in the next generation of microprocessors based on Intel's 80386 which relied on the interaction between microprocessor and software technology while IBM's contribution to system integration was effectively bypassed by many imitators. IBM's efforts to retake its dominant position with new system architecture with the Micro Channel Bus for connecting peripheral devices such as display monitors and communication ports, failed. At present, the future of Intel's position is subject to uncertainty as Microsoft may choose to develop operating system software that allows other IC companies to devise competing microprocessors and as other IC companies reduce the duration of Intel's first mover advantage. Similarly, although Microsoft's position is reinforced by substantial scale advantages, it faces growing competition from other software companies that have benefitted from the involuntary spillovers that have been generated in the broad diffusion of personal computer systems.

The competition among vertical levels for innovative and commercial leadership has clearly been of enormous benefit to the users of personal computers who have seen steady declines in price per unit of performance of systems available in the market. The example of the personal computer is perhaps the most dramatic in illustrating the role of "involuntary" spillovers in encouraging competition and entry in "upstream" industries, but it has been reproduced, on a smaller scale, in many related information technology markets such as mass storage devices, printers, and display technology. This example

suggests the potential value of encouraging competition despite the apparent gains available from scale advantage.

This discussion of the implications of infrastructure for market structure indicates provides a basis for understanding the competitive intensity of information technology industries despite the interests of individual firms in building and sustaining market power. The most important features of the information technology industries are the continuing rapid incremental advance of enabling technologies (e.g. integrated circuit, magnetic storage, and optical) and the potential for entry that can be simultaneously imitative (creating products that can be recognized as similar to those already serving useful purposes) and progressive (creating products that improve upon the performance or capabilities of existing products). In response to these features of the enabling technologies almost all firms in the information technology industry have pursued a similar two-part strategy. The first part of the strategy involves either being among the leaders in introducing new products or being a "progressive" imitator; both outcomes produce temporary quasi-rents that encourage further entry and competition. The second part of the strategy employs compatibility with existing systems and encouragement for the development of complementary products that will enlarge the market and prolong the life of the innovation before its competitive advantage is dissipated through imitation. Ideally, the second part of the strategy which contributes to persistence of an innovation and thus its attractiveness for competitive imitation, will be reinforced by positive externalities originating in the creation of complements and the widespread adoption of the innovation that stimulates more widespread adoption.³² This two part strategy is a direct response to the involuntary spillovers that characterize the technological infrastructure of important sectors the information technology industry. Without these spillovers, both

³² The positive externalities generated by widespread adoption include the creation of human capital that is specific to a particular innovation, the reduction of uncertainty of potential adopters (who can observe that many others have chosen the solution and thus infer that it much be workable), and, for hardware, the likely reduction of input costs as suppliers compete for the larger market. All of these externalities contribute to the persistence of a technology and thus to its infrastructural features when the creator is unable to capture all of the returns from this process.

industry structure and competitive strategy would be different in the information technology industries.

Throughout this discussion the part of technological infrastructure based upon abstract knowledge has been in shadow, existing only in the background of a dynamic process that was largely determined by the innovative activities of commercial enterprises and the involuntary spillovers of knowledge that occur in the course of these processes. This emphasis has been intentional. While it is possible to identify very specific contributions of publicly funded research to the improvement of integrated circuits, software, and even personal computer systems, much of the technological infrastructure in the information technology industry has emerged from commercial development activities.

Is Information Technology Industry Experience Applicable to Public Policy or to Other Industries?

This chapter has stressed the role of involuntary spillovers between enterprises in the information technology industry as a fundamental source of technological infrastructure. Such involuntary spillovers may occur at a very practical level such as our example of the experience with data reduction as an impetus to incremental technological change which is closely related to an economic view of the relation between learning and technological progress. Particularities have played a large role in our description of the operation of the technological infrastructure of the information technology industry. Nonetheless, it is such particularities that establish the practical foundations for commercial progress in the industry.

Several public policy conclusions follow from accepting that "involuntary spillovers" are an important characteristic in the information technology industry technological and competitive performance. It seems straightforward to conclude that public policies encouraging such spillovers are desirable. Alternatively, one may not know how to effectively increase such spillovers but it may be possible to prevent their reduction. There are, however, two problems with this conclusion. First, interventions that seek to preserve such spillovers are equivalent to policies that would prevent the extension of private property rights in the development of knowledge despite the role of such property rights in providing an incentive to create knowledge. Here, there is no easy solution. Existing systems of intellectual property rights were shaped by the dominant technologies of earlier eras and it is likely that reforms that expanded such rights in the area of information technology would stimulate research undertakings. Moreover, it will be unclear *ex ante* whether any particular reform will substantially interfere with the involuntary spillover characterizing the information technology industry. The variety and frequency of both incremental and major innovations in this industry suggest that protection is likely to be short-lived, even if it is extended beyond its current boundaries. Nonetheless, a deeper recognition of the role that involuntary spillovers have played in technological progress in this industry suggests a cautious approach in extending such intellectual property protection.

Second, the ubiquity of increasing returns in the information technology industries recurrently suggests the value of limiting spillovers and entry to tap unexploited economies of scale. Here, the answer is more straightforward. The extent of unexploited returns in the information technology industries is a principal incentive for innovation. It functions best, however, when it remains as a potential rather than actual reward. The ability to leverage the gains found in a dominant position into the control of technological advance and the suppression of variety are the costs that offset the potential gains from tapping the unexploited economies of scale in the information technology industry. While the day may come when fundamental technological limits suggest a rationalization and simplification of the many redundant production and innovation efforts underway in this industry, it has not yet arrived. So long as the rate of cost reductions and performance improvements continue in this industry it would be enormously hazardous to intervene in ways that might discourage the very mechanisms responsible for this progress.

Caution is in order in applying the sort of information technology industry experience described here to other industries. The common infrastructure developed through the pervasive use of computer systems with independently functioning subsystems, the role of digitization in enabling market coordination of diverse IC producers, or the impact of common approaches to the design of personal computers may be relatively rare and fortuitous outcomes in specific segments of the information technology industries. Moreover, as noted in the introduction, the fact that these applications have been enabled by a relatively few materials-related technologies having enormous potential for performance improvement and breadth of applicability are likely to distinguish the experience in information technology from other industries. Nonetheless, this discussion provides a useful starting point for examining the infrastructure of other industries where systems are constructed through the use of components and sub-systems united by common interface methods and sharing underlying technologies. Such industries include a broad range of capital equipment industries in materials, transportation equipment, and aerospace.

By using the type of information technology industry experience described here as a guide, it is possible to examine the extent and consequences of involuntary spillovers in

these other industries as well as the current limits to such spillovers. These limits are one basis for technology policy. Policies that would encourage variety and dis-integration must be weighed against the value of scale economies and design integration. The complementarities between the public good features of publicly funded knowledge creation and the involuntary knowledge spillovers from commercial activities suggest that there are no "neutral" technology policies with regard to technological infrastructure.

References

Abramovitz [1989]

Moses Abramovitz, *Thinking About Growth and Other Essays on Economic Growth and Welfare*, Cambridge: Cambridge University Press.

Aiken [1956]

Howard Aiken, "The Future of Automatic Computing Machines," in *Elektronische Rechenmaschinen und Informationsverarbeitung*, proceedings of a symposium published in *Nachrichtentechnische Fachberichte* no. 4, Braunschweig: Vieweg, pp. 32-34.

Barabaschi [1992]

Sergio Barabaschi, "Managing the Growth of Technical Information" in Nathan Rosenberg, Ralph Landau, and David C. Mowery, *Technology and the Wealth of Nations*, Stanford: Stanford University Press, pp. 407-434.

Blakeslee [1975]

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

Brand [1983]

Stewart Brand, *The Media Lab: Inventing the Future at MIT*, New York: Viking.

Braun and Macdonald [1982]

Ernest Braun and Stuart Macdonald, *Revolution in Miniature: The History and Impact of Semiconductor Electronics*, Cambridge: Cambridge University Press.

Bush [1945]

Vannevar Bush, *Science: The Endless Frontier*, United States Office of Research and Development, Washington, D.C.: U.S National Science Foundation (reprint 1960).

Ceruzzi [1986]

Paul Ceruzzi, "An Unforeseen Revolution: Computers and Expectations, 1935-1985" in Joseph Corn (ed.), *Imagining Tomorrow*, Cambridge: MIT Press, pp. 188-201.

Cohen and Levinthal [1990]

W.M. Cohen and D.A. Levinthal, "Absorptive Capacity: A New Perspective on Learning and Innovation," *Administrative Sciences Quarterly* 35, pp. 128-152.

Cohen and Zysman [1987]

Stephen Cohen and John Zysman, *Manufacturing Matters*, New York: Basic Books.

David [1975]

Paul A. David, *Technical Choice, Innovation and Economic Growth*, Cambridge: Cambridge University Press.

David and Foray [1994]

Paul A. David and Dominique Foray, "Accessing and Expanding the Science and Technology Knowledge-Base: A Conceptual Framework for Comparing National Profiles in Systems of Learning and Innovation," Working Group on Innovation and Technology Policy, OECD, DSTI/STP/TIP(94)4, Paris, April.

David and Steinmueller [1994]

Paul A. David and W. Edward Steinmueller, "Economics of Compatibility Standards and Competition in Telecommunications Networks," *Information Economics and Policy* 6, pp. 217-241.

Flamm [1987]

Kenneth Flamm, *Targeting the Computer: Government Support and International Competition*, Washington, D.C.: The Brookings Institution.

Flamm [1988]

Kenneth Flamm, *Creating the Computer: Government, Industry, and High Technology*, Washington, D.C.: The Brookings Institution.

Kuznets [1966]

Simon Kuznets, *Modern Economic Growth*, New Haven: Yale University Press.

McGraw [1984]

Thomas K. McGraw, *Prophets of Regulation*, Cambridge: Belknap Press of Harvard University Press.

van Meijl [1994]

Hans van Meijl, "The Influence of Spillovers, Product Differentiation, and Entry on Technological Change," MERIT Research Memorandum, 94-026, Maastricht, The Netherlands.

Mowery and Steinmueller [1994]

David C. Mowery and W. Edward Steinmueller, "Prospects for Entry by Developing Countries into the Global Integrated Circuit Industry: Lessons from the United States, Japan, and the NIEs, 1955-1990" in David C. Mowery, *Science and Technology Policy in Interdependent Economies*, Boston: Kluwer Academic Publishers.

Nelson and Winter [1982]

Richard R. Nelson and Sydney Winter, *An Evolutionary Theory of Economic Change*, Cambridge: Harvard University Press.

Okimoto, Sugano, and Weinstein [1984]

Daniel I. Okimoto, Takuo Sugano, and Franklin B. Weinstein (eds.) *Competitive Edge: The Semiconductor Industry in the U.S. and Japan*, Stanford: Stanford University Press.

Prahalad and Hamel [1990]

C.K. Prahalad and Gary Hamel, "The Core Competence of the Corporation," *Harvard Business Review* 90, May-June, pp. 79-91.

Pugh, Johnson, and Palmer [1991]

Emerson W. Pugh, Lyle R. Johnson, and John H. Palmer, *IBM's 360 and Early 370 Systems*, Cambridge: MIT Press.

Ritchie [1986]

David Ritchie, *The Computer Pioneers*, New York: Simon and Schuster.

Rosenberg [1976]

Nathan Rosenberg, *Perspectives on Technology*, Cambridge: Cambridge University Press.

Rosenberg [1990]

Nathan Rosenberg, "Why do Firms do Basic Research (With their Own Money?)," *Research Policy* 19(2), April, pp. 165-174.

Steinmueller [1988]

W. Edward Steinmueller, "Industry Structure and Government Policies in the U.S. and Japanese Integrated-Circuit Industries" in John B. Shoven, *Government Policy Towards Industry in the United States and Japan*, Cambridge: Cambridge University Press, pp. 319-354.

Tassey [1992]

Gregory Tassey, *Technology Infrastructure and Competitive Position*, Norwell, Massachusetts: Kluwer Academic Publishing.

Teece [1989]

David J. Teece, "Profiting from Technological Innovation: Implications for Integration, Collaboration, Licensing and Public Policy" in David J. Teece (ed.), *The Competitive Challenge: Strategies for Industrial Innovation and Renewal*, Cambridge: Ballinger, pp. 185-219.

Williams [1985]

Michael R. Williams, *A History of Computing Technology*, Englewood Cliffs, New Jersey: Prentice-Hall, Inc.

Winter [1989]

Sidney G. Winter, "Knowledge and Competence as Strategic Assets" in David J. Teece (ed.), *The Competitive Challenge: Strategies for Industrial Innovation and Renewal*, Cambridge: Ballinger, pp. 159-184.