Strategic Management Journal
Strat. Mgmt. J., **28**: 563–584 (2007)
Published online 16 March 2007 in Wiley InterScience (www.interscience.wiley.com) DOI: 10.1002/smj.622
Received 3 November 2003; *Final revision received 3 January 2007*

ALLOCATION OF INVENTIVE EFFORT IN COMPLEX PRODUCT SYSTEMS

SENDIL K. ETHIRAJ*
Stephen M. Ross School of Business, University of Michigan, Ann Arbor, Michigan, U.S.A.

This paper examines the allocation of inventive effort in complex product systems. I argue that complex product systems, e.g., personal computers (PCs), are distinguished by functional interaction among several components, each guided by a relatively autonomous bundle of technical and economic characteristics. I try to explore whether the dynamics of such interactions between components of complex product systems can help us understand changes in the relative allocation of inventive effort. I advance and empirically test three hypotheses: (1) emergence of component constraints (bottlenecks) in product systems will trigger research and development (R&D) investment to resolve the constraints; (2) slack component firms have a strong incentive to invest in resolving component constraints; and (3) the incentive of slack component firms to invest in resolving component constraints is increasing in their prior sunk R&D investments in slack components. In sum, I argue that interactions between components in a product system conditions the R&D incentives of firms and also that the incentives are increasing in their prior investments or capabilities. Using product reviews from technical journals, I trace the constraint components in the PC from 1981 to 1998 and attempt to predict shifts in the allocation of inventive effort in the subsequent period. The empirical results strongly support all three hypotheses. This study highlights the paradoxical effect of modularity in complex product systems. Modular design architectures, while contributing to accelerating the pace of technical change, also tend to limit the economic benefits of firms' component R&D efforts, especially when different components technologies are progressing at different rates. This often creates an impetus to enlarge the scope of firm R&D activities beyond the component product markets that firms operate in. Other implications for R&D decision making are discussed. Copyright © 2007 John Wiley & Sons, Ltd.

INTRODUCTION

The problem of designing and coordinating the activities of large-scale complex systems, such as products or organizations, is central to the management research enterprise. Simon (1962), in his seminal essay on the *Architecture of Complexity*, suggested that systems that are hierarchical and nearly decomposable (i.e., modular) help mitigate the effects of complexity. More recent work advocating the use of modular design principles in complex products (Baldwin and Clark, 2000) and organizations (Sanchez and Mahoney, 1996) has resurrected interest in modularity as a vehicle to manage complexity. In the product design context, modular design structures are favored over their integrated counterparts when design objectives such as flexibility and rapid innovation are more important than overall product performance (Ulrich and

Keywords: inventive effort; R&D scope; complex systems; modularity
*Correspondence to: Sendil K. Ethiraj, Stephen M. Ross School of Business, University of Michigan, 701 Tappan Ave., ER 4608, Ann Arbor, MI 48109, U.S.A. E-mail: sendil@umich.edu

Eppinger, 1999). Modularity accelerates product innovation by allowing increased specialization of knowledge, parallel and autonomous innovation opportunities in different modules, and the recombination of modules to create superior system configurations, respectively (Baldwin and Clark, 2000; Schilling, 2000). Thus, the central implication is that using modular designs in complex product systems reduces the scope of research that firms need to pursue in-house, relying instead on the market for noncritical, but complementary, components (Schilling and Steensma, 2001).

The present paper examines the allocation of research and development (R&D) effort in modular, complex product systems. I argue that complex product systems (e.g., personal computers) are distinguished by interaction among several components, each guided by a relatively autonomous bundle of technical and economic characteristics. Thus, if the interactions between components are public and well documented, R&D and innovation in the components can occur relatively independently of one another. There is, however, a blemish in this picture of how modularity ushers specialization in the targets of innovation efforts. Since product performance depends significantly on how the components functionally interact, innovation in one or more components can alter the nature and magnitude of interaction between components. Such changes in the interactions between components can alter the incentives of firms to specialize in their respective components, and even shift R&D efforts into components that they do not produce.

The basic intuition that I try to explore is whether the dynamics of interactions between components of complex product systems can help us understand changes in the allocation of inventive effort between them. I build on the idea that there is an induced component of inventive effort that is triggered by the emergence of component constraints to system performance (Rosenberg, 1969). These component constraints alter the marginal returns to inventive effort in component technologies, and thus guide inventive effort in particular directions. This paper offers three main contributions. First, the paper elaborates how the dynamics of component-level interactions partially shape the technological trajectories in complex product systems and subtly guide firm R&D incentives in particular directions. Second, the paper highlights a paradoxical effect of modularity in complex product systems. Modular design architectures, while contributing to accelerating the pace of technical change, also tend to limit the economic benefits of firms' component R&D efforts especially when different component technologies progress at different rates. This often creates an impetus to enlarge the scope of firm R&D activities beyond the component product markets that they participate in. Thus, contrary to the expectation that modularity promotes specialization and the consequent reduction in firm R&D scope, I hypothesize and empirically demonstrate that, under certain conditions, on average, all firms in the product system, regardless of the component(s) they produce, have an incentive to resolve component constraints. Lastly, I complement the existing literature that has empirically documented that firms' R&D scope is broader than their product scope (Patel and Pavitt, 1997) and sought explanations for why this is so (Brusoni, Prencipe, and Pavitt, 2001). Whereas Brusoni *et al.* (2001) focus on systems integration ability as the primary motive for such an expansion in the scope of R&D activity, I argue and show that R&D scope widening is a characteristic not limited to systems integrators (e.g., Dell in personal computers, Hewlett-Packard and Sun in workstations, Boeing and Embraer in aircraft) but extends also to component producers in complex product systems. The primary motive for the component producers is to release and exploit the value of cumulative component R&D investments.

Using data from the personal computer (PC) product system, I show that constraint components—the component(s) that poses the greatest bottleneck to improving system performance—are an important predictor of allocation of inventive effort in the subsequent period. The results indicate that the identification of a component as a constraint to system performance explains a 13 percent shift in the allocation of total inventive effort to that component. I also show that the slack component (the component that does not pose a bottleneck to system performance) firms, on average, shift about 8.5 percent of their R&D investment in an apparent attempt at resolving such component constraints. Finally, the incentive of slack component firms is increasing in their prior sunk R&D investments. A 1 percent increase in prior R&D stock in slack components contributes to a 0.8 percent increase in investment to lift component constraints.

DOI: 10.1002/smj

The rest of the paper is organized as follows. In the next section, I elaborate on the notion of complex product systems and develop the theory and hypotheses. This is followed by the description of the study setting and the research design. I then present the results and discuss their implications.

THE DYNAMICS OF INVENTION IN COMPLEX PRODUCT SYSTEMS

Simon (1962: 468) defined a complex system as 'one made up of a large number of parts that interact in a nonsimple way ... [in] such systems ... given the properties of the parts and the laws of their interaction, it is not a trivial matter to infer the properties of the whole.' The complexity here stems primarily from the often unknown nature and magnitude of interactions between different parts of the system and the consequent system performance implications. For instance, the nature of interaction between two parts may range from positive (increasing in one another), negative (decreasing in one another), or unrelated. Furthermore, the nature of the interaction may alternate between positive, negative, and unrelated over different magnitudes of the interaction relationship. As a result, overall system performance can exhibit highly nonlinear and/or nonmonotonic behavior in response to changes in one or more parts. A complex product system is, therefore, a special case of a complex system and may be usefully viewed as a set of components that together provide utility to customers (Garud and Kumaraswamy, 1995: 94). In other words, the value of the components together is greater than the value separately.

Modularity is a general set of design principles for managing the complexity of such large-scale product systems. It involves breaking up the system into discrete chunks that communicate with each other through standardized interfaces or rules and specifications. Baldwin and Clark (2000: 63) define a module as 'a unit whose structural elements are powerfully connected among themselves and relatively weakly connected to elements in other units.' Thus in a complex product system with a modular architecture, we can conceive of at least three levels of analysis: the system, the components, and the firms that design and manufacture the components. Interactions within and across the three levels are the key loci of change and innovation. I examine each in turn.

System level

A distinguishing characteristic of complex product systems is that product or system performance is in part a function of how components interact together as a system (Garud and Kumaraswamy, 1993; Prencipe, 1997). The interaction relationship among the components of the product also determines the marginal contributions of the various components to product performance (Hoetker, 2006; Ulrich, 1995). For instance, a car engine may be rated at 100 brake horsepower (BHP). However, the realized BHP may be determined, among other things, by the interaction of the engine with the weight of the chassis, ergonomics of the car body, the fuel injection system, and so on (Clark and Fujimoto, 1991: Ch. 5). Because of this characteristic of complex product systems, even when the system is operating at peak performance not all its components are likely to be operating at their highest rated capacity as well (Sahal, 1981: 29). While some may be operating at their highest rated capacity, others may still have some slack. This means that changing the parameters of one component of the product can have fairly large and unpredictable consequences for product performance simply because the interaction relationship among the components of the product is altered (Ethiraj and Levinthal, 2004b). At the extreme, such micro changes to parts of the product can result in highly dysfunctional effects on product performance.

An illustration is useful in making the argument clearer. For instance, in a PC, the random access memory (RAM) unit loads data (or input) and transfers them via the system bus to the microprocessor for processing. The microprocessor passes the results back to the RAM unit after processing. System performance is the time taken for one cycle of the task to be completed. We can see that system performance (cycle time) can be improved by increasing the speed and/or processing capacity of the two components or changing the interface (i.e., system bus) between them. On the one hand, if one of the components (say RAM) is performing at peak capacity and the second component is operating below capacity, system performance can be improved only by innovating in the former. On the other hand, if RAM is kept constant and the microprocessor's capacity is improved, then PC performance can actually degrade if the capacity

DOI: 10.1002/smj

and speed of the RAM are significantly less than the demands of the microprocessor.

Fixing these ideas more formally, consider a single-product firm (i.e., a firm that competes in a single-component product market), F, whose profits (*ceteris paribus*), Π, are a function of its accumulated R&D stock, R, and its annual flows of R&D, r. Let P denote system product performance and p component product performance. Thus, we define firm profits as

$$\Pi(P(R, r)) = \beta(p(R, r)) - C(r) \qquad (1)$$

Firm profit in Equation 1 is a function of how R&D stock and flows are deployed to improve system performance. $\beta(p(R, r))$ is the firm's return from doing R&D to improve component performance, p, and $C(r)$ is the total cost of R&D flows in each period. The key idea is that firms only engage in component R&D and their returns are a function of how component R&D, r, affects component performance, p, and subsequently system performance, P.

Let p_s denote the performance of the slack component(s), and p_c the performance of the constraint component(s). Then by definition, improvement in slack component performance yields no improvement in system performance and sometimes even degrades performance because the constraint component limits the utilization of slack components. Conversely, improvement in the constraint component yields improvement in system performance. Stated more formally:

$$\frac{\partial P}{\partial p_c} > 0, \text{ and } \frac{\partial P}{\partial p_s} \leq 0 \qquad (2)$$

Further, as I argued earlier, in a complex system, system performance is a function of interdependencies among the components (Ulrich, 1995). In other words, unconstrained improvement in one or more components yields no system performance improvement since the interdependencies among the components limits the realization of system performance improvement (see Rosenberg, 1969, 1974, for several examples of imbalanced technical change). Thus, system performance is a function of the interaction among the various components (Shibata, Yano, and Kodama, 2005). If p_s and p_c represent the set of all slack and constraint components, respectively, then changes in system performance are a function of the changes in the interaction among the components. Stated more formally:

$$\frac{\partial P}{\partial p_s \partial p_c} \neq 0 \qquad (3)$$

Combining Equations 2 and 3 above suggests that system performance, P, changes as the interaction among components change, and that system performance improves with improvement in constraint component performance, p_c. Earlier, I argued that improvement in system performance is a function of component performance improvement. If system performance improvement occurs as a function of constraint component improvement and system performance improvement drives profits, then it follows that returns to R&D in constraint components must be greater than zero and immediate returns to R&D in slack components must be zero or negative. Stated more formally:

$$\frac{\partial \Pi}{\partial r_c} > 0 \text{ and } \frac{\partial \Pi}{\partial r_s} \leq 0 \qquad (4)$$

where r_c is the annual flow of R&D investment in the constrained component(s) and r_s the annual flow of R&D investment in the slack component(s).

If returns to R&D in constraint (slack) components are greater (equal to or less) than zero, then a firm facing a choice of investing R&D in constraint or slack component is, *ceteris paribus*, likely to have a greater incentive to invest in R&D in the constraint component. Stated more formally:

Hypothesis 1: $\frac{\partial \Pi}{\partial r_c} > \frac{\partial \Pi}{\partial r_s}$

Hypothesis 1 states that, across a complex product system, the returns from R&D directed at the constraint component will outstrip the returns from R&D to the slack components. Thus, an empirical test of this hypothesis will involve the demonstration of a shift in R&D effort devoted to constrained component(s) relative to the slack component(s). In summary, the discussion above suggests that if we assume system performance to be a function of components' performance and, in particular, that it may be a min function of the performance of the various components, then: (1) component innovation is necessary when one or more component technologies are a constraint[1] to product performance; and (2) marginal returns to R&D directed at removing constraints is significantly positive, whereas marginal returns to R&D in slack components is zero or negative.

Component interactions

I argued above that the performance of complex product systems is a min function of components' performance and that one route to performance improvement is innovations directed at component constraints. Innovations that lift component constraints, in addition to helping improve system performance, tend to alter the interaction relationships between components in the product system (Rosenberg, 1969). This, in turn, alters the marginal performance contribution of the components. Thus, resolution of one or more component constraints in one period can result in the emergence of one or more other components as constraints in the following period.

The point may be illustrated with an example. Assume that returns to R&D in PCs are highly correlated with computing speed measured as MIPS (million instructions per second) or MFLOPS (million floating point operations per second). Therefore, the R&D objective of PC component design is to maximize MIPS or MFLOPS. Realized MIPS or MFLOPS will depend on: (1) the processing power of the microprocessor; (2) the amount of memory in the system; (3) the width of the pipeline (bus) between the microprocessor and the memory; (4) the speed of physical disk drive; and (5) the size of the data on which the computation is to be performed. If the width of the pipeline bus becomes a constraint to system performance, no improvement in memory or processing power of the microprocessor will make any difference to system performance. An optimal performance given the constraints is reached when no further improvements are possible without changing one or more of the constraint components' parameters. Once the constraint in the pipeline bus is lifted, some of the slack in one or more other components is absorbed and system performance improves. At the same time, one or more other components become constraints to system performance, thus increasing the marginal returns to innovation in those components. The key observation of this paper is that the period-to-period observed component constraints guide the R&D agenda of firms by altering marginal returns to R&D effort in each component.

The stylized dynamics of innovation in complex product systems suggests that following innovation in a component that is constraining system performance we can expect to observe: (1) increase in system performance; (2) changes in the marginal performance contribution of different components; and (3) different component(s) emerging as a constraint to system performance.

Firm level

At the firm level, the issue of who invests in resolving component constraints pertains to the issue of firm-level R&D scope or technological competencies and its performance implications (Griliches, 1984; Nelson, 1959, 1961; Pavitt, Robson, and Townsend, 1989). Investigating this empirically, Patel and Pavitt (1997: 144–145) found that large firms typically invest in a much broader range of technical fields than the core product fields in which they compete. For instance, only 71 percent of chemical firms' R&D is in chemical technologies. The remaining R&D spans nonelectrical machinery and electrical/electronics technologies. Similarly, only about 77 percent of computer industry R&D is concentrated in the electrical/electronic industries. They offer two reasons for such technological diversity (Patel and Pavitt, 1997). First, the technological interdependence between manufacturers of products and the components manufactured by suppliers requires complex coordination on product improvements, thus creating the imperative to cross technological boundaries. Second, exploiting emerging opportunities often requires firms to straddle new technological domains, again contributing to an increase in the size of firms' R&D portfolios (Patel and Pavitt, 1997: 148). In other words, the primary impetus for technological diversity lies along the vertical value chain.

In more recent work, Pavitt and his colleagues (Brusoni *et al.*, 2001), focusing on systems integrators, return to the question of why firms know more than they make. Using data from the aircraft engine control industry, they suggest that firms

[1] It is also possible that the interface, not the component(s), is the constraint. In such cases, architectural, rather than component, innovation is necessitated. This paper does not deal with architectural innovation.

DOI: 10.1002/smj

need to know more than they make to cope with the imbalances caused by uneven rates of development in the technologies embedded in products (Rosenberg, 1969) and unpredictable product interdependencies (Ulrich and Eppinger, 1999). Their work, while enriching our understanding of R&D scope choices made by systems integrators such as Boeing, Airbus, or Dell, provide little guidance on whether (and how) similar mechanisms might explain the R&D behavior of component firms as well.

A complex product system (say PCs) usually includes several firms with heterogeneously distributed capabilities and operating in one or more distinct component industries. Each firm faces an R&D budget constraint and the problem is one of allocating the budget to maximize returns from R&D. However, since firms are likely to specialize in one (or few) component(s), the incentives to invest in resolving component constraints are unlikely to be uniform across firms. In examining firm incentives to invest in removing component constraints, it is useful to distinguish between two sets of firms: (1) C-firms (F_c), firms that compete in the product market of a component that is a constraint in the current period; and, (2) S-firms (F_s), firms that compete in the product market of slack components. As I argue below, even the S-firms have a strong incentive to resolve constraints.

S-firms

The S-firms are firms that design and/or manufacture slack or nonconstraint components. It would appear that the S-firms have little incentive to allocate R&D resources to resolve component constraints since they would be unable to directly appropriate any returns from R&D allocated to resolving the constraints. However, the unique characteristic of complex product systems—interdependencies between components—creates a strong incentive for even the S-firms to allocate R&D resources to resolving component constraints. The existence of component constraints not only reduces the incentive for slack firms to invest annual R&D flows in slack components, but also reduces the ability of the S-firms to appropriate returns from sunk R&D investments in slack components. The residual value of slack component R&D stock can be freed only by resolving the component constraint. Thus, I argue that S-firms have a strong incentive to direct their annual R&D flows to resolve component constraints even if they are not producing the constraint component(s).

We know that not all firms have the same incentive to invest in lifting constraints. The incentive to invest is conditioned largely by the appropriability regime that firms operate in (Arrow, 1962). If firms believe that they can capture a significant proportion of the returns from R&D, then they are likely to invest. However, if the benefits of R&D are like a public good, i.e., the inventing firm cannot exclude other firms (e.g., due to poor intellectual property protection) from exploiting its R&D, then there is likely to be underinvestment. In such cases, there is likely to be a free rider problem with firms waiting for others to resolve the constraint. In other words, the cost of investing first outstrips expected economic returns, causing firms to wait and watch so that they can free ride on other firms' investments to lift the constraints.

The free rider problem disappears, however, when there is a significant cost associated with waiting. For instance, if the economic returns to R&D are concentrated in a short time window, then the firm that moves first will capture a large portion of the rents before the imitating firms can catch up and erode the advantage of the innovator. There is a strong incentive here for firms to not only invest quickly, but also be the first to come up with an innovation that lifts the constraint. This provides the innovating firm with the short window to skim the economic rents before imitators compete away the value. In this case, one imitating firm has few advantages over other imitators and free riding has little strategic value. The 'racing to invest' behavior is often observed in high-technology industries such as computers, telecommunications, and video games (Dasgupta, 1988; Khanna, 1995). The distinguishing feature of such regimes is a short product life cycle where the innovator has a small time window to reap the value of an invention. For instance, in disk drives, where product life cycles range between 1 and 2 years (Christensen, 1997), if firms miss one generation their subsequent survival chances are greatly reduced. Similarly, in the computer industry 'Moore's law[2] has been a technological beacon to firms in the field' (Patel and Pavitt, 1997: 154).

[2] Moore's law refers to the now famous miniaturization hypothesis of Gordon Moore, one of the founders of Intel Corp., that the number of transistors on a chip would double every 18 months.

DOI: 10.1002/smj

Thus, I expect that firms will have a strong incentive to invest in resolving the constraint (rather than wait and watch) when the returns to R&D are concentrated in a short time window.

There is ample anecdotal evidence in contemporary R&D activity that supports this argument. For instance, Intel developed the PCI bus back in 1992 as a response to the constraint of the ISA bus introduced in 1984. Even though Intel was not competing in the bus architecture market, it developed the architecture and licensed it to the industry without royalties. The following quote drawn from Intel's web site captures the spirit of how constraints affect R&D effort and the motivation for it:

> PCI was originally developed through the efforts of Intel and other industry leaders in response to the realization that ISA, a bus design developed in 1984, was becoming a bottleneck within the computer. Intel spurred the industry by forming a PCI special interest group with others in the industry, contributing technology from Intel research and development. The result was PCI—one of the most successful chip and board interconnect technologies in history. (Intel Corp., 2006)

The PCI architecture that Intel introduced was followed up with the PCI Express architecture in 2004 that Intel again spearheaded and sponsored. Similarly, Intel introduced the USB architecture for interconnection of peripherals for the PC. Intel does not compete in the peripherals market and will not be able to directly appropriate the returns from the USB technology investment. However, why Intel chooses to do so is seen from the following quote from Bala Cadambi, the Manager for the PCI Express initiative:

> Intel is in the business of providing the engine for the PC, just like Honda is in the business of providing the engine for the automobile. That [PC] engine is doubling in capacity every 18 to 24 months—that's Moore's law. What we really want is to ensure the rest of the platform goes with it. This means that if the engine gets better, the tires get better, the chassis gets better, the roads get better, and you get better gas mileage ... we want the platform, which is everything around the microprocessor, to be keeping pace and improving and scaling, so the microprocessor can deliver its potential. (Cadambi, 2006)

Similarly, AMD, a chip manufacturer, is investing in R&D in security technologies. AMD's investment in security technologies appears spurred by possible bottlenecks in security technologies that may be constraining the adoption and growth of computers, though it does not produce security software or compete in the security software industry. AMD has an incentive to invest in security technologies to the extent that PC security concerns dampen the sales of AMD chips. The following quote from AMD captures the importance of security and how it guides their R&D effort:

> AMD is working closely with its chipset ecosystem and key operating system partners to bring a united security solution to market. This solution will coordinate next-generation AMD processor features for secure virtualization, new chipset features for secure input and output, and new platform components mapping to trusted computing requirements, such as the TPM. (AMD, 2006)

AMD also invested in a graphics interface standard to resolve graphics bottlenecks in the PC. The following quote from an AMD white paper clearly illustrates how bottlenecks guide and shape R&D investments:

> Tremendous efforts have been made to improve computing system performance through increases in CPU processing power; however, I/O bottlenecks throughout the computing platform can limit the system performance. To eliminate system bottlenecks, AMD has been working with industry leaders to implement innovative technologies including AGP, DDR SDRAM, USB2.0, and HyperTransport™ technology ... The Accelerated Graphics Port (AGP) was developed as a high-performance graphics interface. It alleviates the PCI graphics bottleneck by providing high bandwidth throughput and direct access to system memory. (Chen, Johnson, and Suhrstedt, 2002)

Finally, Seagate Technologies, a disk drive manufacturer, invested in the serial ATA standard in response to the constraint in data communication between the system board and the disk drive. In all of the examples cited, the companies did not seek to directly appropriate the R&D investments in resolving the component constraints. They licensed these technologies to the industry under royalty-free terms so as to promote their diffusion and thus lift the constraints.

Fixing these ideas more formally, consider two single-product firms; F_s, which produces a slack component product, and F_c, which produces a constraint component. Each firm has stocks of R&D in the slack component, R_s, and the constraint component, R_c. Therefore, by Equations 1, 2, and 3

DOI: 10.1002/smj

the returns from cumulative R&D in the constraint component(s), R_c, is positive and, conversely, the returns from cumulative R&D in the slack component, R_s, is zero or negative. Thus, Equation 4:

$$\frac{\partial \Pi}{\partial R_c} > 0 \text{ and } \frac{\partial \Pi}{\partial R_s} \leq 0$$

In other words, the slack component firms face depressed incentives to add annual R&D flows, r_s, to the slack component. In order to extract value from their accumulated stock of slack R&D, R_s, the constraint has to be lifted. Thus, slack component firms have an incentive to invest annual R&D flows in resolving component constraints that will release opportunities to extract the value of accumulated R&D stocks, R_s. In the examples above, Seagate had to invest in improving data interface technology, a constraint component, so that the disk drive investments could be recouped. Thus, for slack firms, the returns to investing R&D flows in constraint component(s) are positive. Put formally:

Hypothesis 2: $$\frac{\partial \Pi^{F_s}}{\partial r_c^{F_s}} > 0$$

where the superscript F_s denotes the slack component firms.

Finally, there is the question of whether all slack component firms have the same incentive to invest in resolving the constraint. From Dierickx and Cool (1989) we know that asset stocks are an important source of firm competitive advantage. It follows, therefore, that firms with the greatest sunk investments in slack component R&D are likely to face the greatest incentive to invest in resolving the constraint. This is because for every marginal dollar of R&D flows invested in the constraint component a firm with large slack component R&D stock will receive a multiple of the R&D flow investment. The lifting of the constraint will help release the latent value of the accumulated sunk investments in slack component R&D. Furthermore, cumulativeness of R&D investment (such as in the semiconductor industry) provides a strong incentive for slack incumbent firms to invest in lifting constraints since there is a strong incumbency advantage associated with such industries (Nelson and Winter, 1982; Winter, 1984). Thus, I expect that the slack component firms' incentive to invest annual R&D flows in the constraint component is an increasing function of the R&D stock, R_s, in the slack component(s). Stated more formally:

Hypothesis 3: $$\frac{\partial \Pi^{F_s}}{\partial R_s^{F_s} \partial r_c^{F_s}} > 0$$

The following section briefly describes the empirical study setting—the PC product system—and why it is an attractive setting to test the three hypotheses explicated above.

PERSONAL COMPUTER TECHNOLOGICAL SYSTEM

The personal computer is a computer designed for use by a single person. The IBM PC, from its first release, used a modular design, by which different functions were performed by distinct components. The components were designed and manufactured by different firms. The design coordination of the components was achieved by standardizing the specifications of the PC architecture. The performance of the PC, however, depends significantly on the way the components functionally interact. A PC comprises several components and its configuration varies depending on the purpose for which it is used. Also, the configuration of the PC has tended to vary over time, with some components disappearing altogether (e.g., floppy disk) and new ones emerging (e.g., optical drives) over the 1981–98 time period. For these reasons, I chose five PC hardware component industries—system board, microprocessor, memory, display adapter, and hard disk drive—to include in this study. These five components are central to the PC and have existed since its creation in 1981. This provides a consistent baseline to examine allocation of R&D effort in one or more of the five components.

- *System board*. The system board is the heart of the PC on which the key components are embedded. It houses the microprocessor and physical memory, and interfaces with all other components (e.g., hard drives, monitor, input devices). Between PC product generations, the system board has mostly evolved in response to the technical needs of other components. Within a product generation, the goal of production cost reduction has guided its evolution.

- *Microprocessor*. The microprocessor is a logic device that performs the processing function in the PC. The logical structure of the microprocessor comprises three parts: the input/output unit (I/O), the control unit, and the arithmetic/logic unit (ALU). The I/O unit links the microprocessor to the rest of the computer. The control unit is a logic circuit that acts as the interface between the I/O unit and the ALU. The ALU interprets the instructions passed by the control unit and executes them. The results are then passed out of the microprocessor through the I/O unit.
- *Memory*. The primary function of memory (RAM) is to interface between the user and the microprocessor. It stores data and instructions (inputs) and passes them on to the microprocessor, and also stores the output to pass back to the user.
- *Display adapter*. The display adapter acts as the interface between the PC and the monitor. Its primary function is to organize the data in the computer to present it visually on the monitor.
- *Hard disk drives*. Hard disk drives are the long-term memory of the computer. They store data, information, and software in a form accessible by the PC. Hard disk drives have evolved on a trajectory of decreasing physical size, larger capacity, and faster speeds (see Christensen, 1997, for a detailed description of hard disk drive technology changes).

A significant limitation of the study is the noninclusion of the PC operating system and important application software. Clearly, software is a critical component of the PC system and its inclusion is warranted. Unfortunately, I was unable to include software component firms in the study because I use patents to decompose the R&D effort of firms into their respective component domains. In the software industry, historically, copyrights rather than patents have been the source of intellectual property protection. This precluded the decomposition of software firms' R&D budgets into component technology domains. While this is clearly a limitation of the present study, it is common to other studies as well (see Patel and Pavitt, 1997: 143). Nevertheless, the important caveat is that the transition from constraint to slack components could still be triggered by software changes that are exogenous to the model.

The PC product system comprises subsystems of the above component products. Each component is subject to its distinct market selection criteria in addition to the selection environment of the PC product system as a whole. Also Moore's law has largely guided the product life cycle of PCs, with just 2–3 years separating product generations (Langlois and Steinmueller, 1999). This suggests that firm R&D investment is unlikely to be driven by free-riding behavior. In addition, R&D investments in the computer industry have tended to be cumulative rather than discrete (Malerba *et al*., 1999). This increases the importance of sunk R&D stocks in driving decisions about annual R&D flows. Thus the PC product system is an attractive setting for testing the three hypotheses outlined earlier.

RESEARCH DESIGN

Sample

The sample for the empirical study was drawn from the Corptech Directory of Technology Companies. The Corptech directory is available each year starting from 1985. Since the PC industry began around 1981, drawing the list of companies from 1985 onwards is only an approximation. As a result of this sampling procedure, firms that entered and exited the industry during the period 1981–85 will be omitted from my dataset.[3] The directory is organized along industry and product lines. I included all firms that design/manufacture the five components (system board, microprocessor, memory, display adapter, and hard disk drives) described above. I identified a total of 173 firms, both public and private, from the United States, Europe, Japan, and Taiwan that belonged to the five component industries. A fortunate coincidence is that almost all the firms in the dataset (with the exception of nine firms) generate at least 80 percent of their sales revenues from a single component product market.[4] Employing the prevailing definitions of firm diversification, nearly all

[3] While left censoring is a clear limitation, it is not very significant in the present study. In ecological studies of entry and exit, left censoring can produce biased estimates of relationships. In the study reported here, the left censoring problem raises some concerns about generalizability (i.e., representativeness of the sample) of the results without affecting the estimation itself.

[4] I ultimately exclude both IBM and Texas Instruments from the analysis as they participate significantly in many different

DOI: 10.1002/smj

firms in our sample would be classified as single-business firms (Rumelt, 1986). Thus, I was able to establish a one-to-one mapping between firms and their component product markets.

The bibliographic information for all the patents granted to the companies in the sample in the period 1979–98 was obtained from the United States Patent and Trademark Office (USPTO). The PC industry began in 1981 with the launch of the first IBM PC.[5] I extracted data from 1979 to estimate 2-year lag models. The data include patent number, title, assignee name, application date, issue date, and original and current patent classes. There is a problem of identifying all the patents granted to a company for a variety of reasons: (1) companies may have been acquired in the 20-year period; (2) companies may have changed names; (3) companies may have filed patents under different names (e.g., AMD and Advanced Micro Devices); and (4) typographical errors made by the USPTO. Such problems are predominant for large firms with several business units filing several hundred patents each year. For this subset of firms, I acquired cleaned data from CHI Research and cross-checked with the Assignee code field that the Office of Technology and Forecasting (OTAF) at the USPTO employs for uniquely identifying patent assignees.

The second task is one of identifying the constraint component(s) in the PC industry during the study period. For this purpose, I examined and analyzed product reviews in *PC Magazine* and *PC World* to identify the key component constraints to system performance. The magazine reviews are available from 1983 onward. Each magazine publishes two issues a month for a total of 24 issues a year. On average, I found about 8–12 detailed component reviews each year that I employed to code the binding component constraint. For instance, consider the following excerpt from a review:

> Over the last few months the glittering image of the PC AT has acquired a certain patina—some would say tarnish—because of problems with the data storage hardware. The core of the problem seems to be in the 20-megabyte hard disk ... Because of their delicate architecture, Winchester-type hard drives have certain inherent limitations. A number of factors, including power failure, shock, and even simple vibration, can produce a head crash that results in partial or total loss of all stored data. More often a crash will occur when the drive head makes contact with a dust particle or some other contaminant on the disk surface. Even if the hardware is not damaged, it's still going to put an end to 10 megabytes or more of files. *(PC Magazine*, June 25, 1985: 126)

The above excerpt conveys at least two pieces of information that allows the identification of a constraint. One, the first line of the excerpt suggests that the performance of the system as a whole (the PC in this case) is affected by the said component. This implies that system performance can be improved by solving the problem with the component. Two, the review also identifies the precise nature of the constraint (i.e., head crash in the hard disk). This provides R&D engineers with a clear problem and thus defines an R&D agenda. These two pieces of information taken together help infer that the hard disk might have been a constraint to effective PC performance in 1985.

For each year I used three to four reviews to triangulate and identify the primary bottleneck(s). This was done to ensure that the bottleneck identified was more likely to be an industry-wide phenomenon rather than peculiar to a company's product or model. In addition, I cross-checked component constraints using reviews from *PC World* to remove magazine-specific variance, if any. Finally, I submitted the set of reviews to three graduate students in Computer Science at the University of Pennsylvania to identify whether they reflect constraints to system performance. The set of components identified as constraints in a year reflect unanimity among the three coders. This was done to remove coder bias, if any. Finally, in the absence of evidence to the contrary, I made the conservative assumption that a bottleneck is resolved within one year. In most cases the assumption was reasonable since new models came up within a year or two following its identification as a constraint. Table 1 lists the years in which a component was identified as a constraint.

The coding approach adopted here is subject to at least one important criticism. The data for coding constraint components came primarily from product reviews published in two PC magazines. It can be argued that the reviews are mostly written for an audience of PC users or consumers

component markets. However, the results do not change even if they are included.

[5] The study is confined to the IBM PC and does not include Apple computers. The value chains for the two are largely distinct and there are few firms that operate in both value chains, making comparisons less justifiable.

Table 1. Component constraints identified in each year

Year	MPU	Memory	Graphics	HDD
1981				
1982				
1983	X			
1984				X
1985			X	X
1986			X	X
1987				
1988	X			
1989		X		
1990			X	
1991			X	
1992				
1993				
1994			X	X
1995		X		X
1996				
1997				

rather than for R&D personnel. This raises concerns about a possible confound in establishing the connection between the reviews and subsequent R&D behavior of firms. While it is true that the reviews are written for consumers, the reviewers themselves are technically competent (e.g., Peter Norton, the creator of Norton Utilities, later acquired by Symantec Corp.) to make accurate and informed evaluations of the components. Moreover, in most cases, the reviews are based on evaluation products that firms themselves make available to the reviewers. Therefore, it seems reasonable to assume that firms might track these reviews and act on them as well.

The study then sought to identify patents related to each of the components, since it is my basic premise that firms operating in complex product systems allocate inventive effort not only in their component products but also in the broader product system. Unfortunately, there is no simple way of classifying patents to product markets since the U.S. patent classification system groups patents by function rather than application. I conducted the assignment process in three stages. First, using patent class descriptions I identified a subset of patent classes clearly associated with the PC industry. This provided a first-pass mapping of patent classes to PC component industries that was generally consistent with a concordance map produced by the USPTO in the late 1970s. Second, I excluded from the dataset any patent subclasses that did not fall within one of the five specified components (e.g., optical drives). Third, I confirmed the classification with industry experts and employees at the Office of Technology and Forecasting and added patent classes that I missed and expanded the classification from three-digit patent classes to six digits in cases where the three-digit classes were too coarse.

Dependent variable: Allocation of inventive effort

I employed patent count data to construct a measure of allocation of inventive effort. First, I obtained patent counts based on patent application dates for each firm year across all the patent classes that constitute the five components. I then computed the ratio of patents in a particular component to the total patents across all components for each firm year, to capture the relative allocation of inventive effort to a particular component. Specifically:

$$Y_{itj} = \frac{\mathrm{PC}_{itj}}{\sum_{n=1}^{N} \mathrm{PC}_{itn}}$$

where Y_{itj} is the allocation of inventive effort by firm i, in year t, in component class j; PC_{itj} is the count of patents filed (application date) by firm i, in year t, in component j; the denominator is the aggregate of patents filed by firm i, in year t, summed over $N(= 5)$ components.

There is considerable debate in the literature over the use of patent-based measures in studies of innovation. It is recognized that firms and industries widely differ in their propensities to patent and also in the motives for patenting (Cohen, Nelson, and Walsh, 2000; Levin *et al.*, 1987). However, the components of the PC system are all subsets of the broader semiconductor industry, and there is a long tradition of research in this area using patent data (see, for example, Almeida and Kogut, 1999, and the studies they cite). There is general consensus that, though imperfect, patents may reliably capture inventive activity in the computer industry (Cohen *et al.*, 2000). Besides, there has been significant criticism only in the use of patents as a measure of innovation (Trajtenberg, 1990), but the general consensus is that patents are a fairly reliable indicator of input-side inventive effort (Hausman, Hall, and Griliches, 1984). For instance, Trajtenberg (1990) in a study of the CT scanner industry found that the correlation between patent counts and R&D ranged between 0.79 and 0.91, whereas patent counts had no statistically significant relationship with independent measures of innovation. This led him to conclude that 'patent counts are good indicators of the inputs to the innovative process' (Trajtenberg, 1990: 179). Consistent with this, the present study uses patents to decompose the research budgets and capture their relative allocation to different components. The primary assumption here is that firms allocate a chunk of research to a product system and then partition this budget among the components. Patenting in the component classes allows us to proxy the relative allocation of firm R&D effort.

Explanatory variables

Constraint component

The constraint component in each year that was identified from the product reviews was coded as a dummy variable taking a value of 1 for the component with a 1-year lag. For example, I found that the microprocessor was listed as a constraint to PC performance in 1983. Taking a 1-year lag, I expect it to influence the relative allocation of inventive effort in the following year, i.e., 1984. The 1-year lag was motivated by two considerations. First, there is anecdotal evidence that the industry has tended to track Moore's law; i.e., technological trajectories follow 18-month cycles. Second, my interviews with R&D engineers in the industry indicated that R&D budget cycles usually lasted 12–14 months. Thus, the results reported here used a 1-year lag. The results were also robust to the use of 2-year lags.

S-firms × constraint component interaction

In addition to examining whether inventive effort in complex product systems responds to component constraints, the study sought to make the case that the slack component firms in the product system have a strong incentive to invest in resolving such constraints. I identified the primary component product market domain of each firm in the sample and created a variable coded 1 if the firm did not produce the component that was identified as a constraint in the given year and 0 otherwise. I interacted this variable with the constraint component dummy to test Hypothesis 2.

S-firms × constraint component × firm slack knowledge stock interaction

Hypothesis 3 argued that the slack component firms' incentive to invest in resolving the constraint is increasing in their sunk R&D investments in the slack component. I created a stock of firm knowledge in the slack component by cumulating the firm stock of patents in the primary slack component and interacted it with the S-firms dummy and the constraint component dummy.

Control variables

Component industry R&D intensity

Several studies document the robust relationship between R&D expenses and patenting (Hausman *et al*., 1984; Trajtenberg, 1990). In order to account for component industry differences in R&D intensity, I measure R&D intensity as a contemporaneous time-varying estimate of the ratio of aggregate R&D expenses in each component to its aggregate sales. The component-level R&D expenses and sales were obtained by aggregating the R&D expenses and sales, respectively, of all public firms listed in Standard & Poor's Compustat in the corresponding four-digit SIC class. It is likely that aggregating the R&D expenses and sales of only public firms understates the actual expenses and sales in the components. However, this is not a significant concern, if the understatement does not vary systematically by component technology. For instance, if mostly private firms populate the system board component industry, then the extent of understatement will be higher. I employ other controls to account for this.

Component industry sales

Schmookler's (1966) central argument is that demand conditions are the basic drivers of allocation of inventive effort. I attempt to account for demand conditions using past sales information. I measure sales as a time-varying estimate of aggregate sales in each component industry lagged by 1 year.

Component industry size

Schumpeter (1950) argued that large firms with R&D programs would account for a disproportionate number of innovations in the economy. The

DOI: 10.1002/smj

R&D–size relationship is explained by a variety of reasons: capital market imperfections that allow large firms to invest greater amounts in R&D; scale economies in research; and better capabilities of large firms to appropriate returns from research. Size is measured as the time-varying estimate of the log of total number of employees in each component industry measured at the four-digit SIC class.

Component industry knowledge base

Prior research indicates that in cumulative technologies the ability to absorb and create new knowledge is conditioned by past knowledge or absorptive capacity (Cohen and Levinthal, 1990). Therefore, patenting propensity in a cumulative technology such as semiconductors is likely to be higher to the extent that there are already a large number of prior patents to build upon. Knowledge base is measured as a time-varying estimate of the cumulative patents (by application date) filed by all firms in the component lagged by 1 year.

Component industry concentration

There is some evidence in the literature that R&D incentives are conditioned by market structure. While Schumpeter (1950) argued that large firms in concentrated industries have a greater incentive for innovation, Arrow (1962) showed that a monopolist has less incentive to innovate than firms in competitive industries. I control for such differences in the five component industries using a time-varying Herfindahl measure of the extent of concentration in each component industry.

Other component industry controls

Patenting propensities and R&D budgets vary systematically by industry. In order to capture such differences, I created four time-invariant dummy variables coded as 1 if the firm belonged to each of the following component industries: (1) microprocessor; (2) memory; (3) display adapter; and (4) system board. The omitted category was hard disk drive component firms.

Firm age

Empirical evidence indicates that organizational age and experience in a particular industry are positively related to patenting (Sorenson and Stuart, 2000). As organizations develop experience with patenting, they develop and refine the routines for search and discovery of incremental inventions that improve their productivity over time. Though the dependent variable here is not patenting, we can expect inventive effort in component technologies to be somewhat related to firm age. As firms age, their routines are likely to get entrenched over time. Such routines may also create inertia in inventive effort and increase consistency with the past (Nelson and Winter, 1982). I measured firm age as the time-invariant estimate of the founding year.

Firm knowledge stock

The cost of patenting is likely to vary significantly across firms. Firms that have patented in the past are likely to face lower costs of patenting in the future as a function of knowledge infrastructure, spillovers from R&D projects, knowledge of the patenting process, and so on. Therefore, I expect past experience in a component technology to increase the patenting propensity of firms. In addition, it is important to control for the main effect of firm knowledge stocks since Hypothesis 3 is based on an interaction effect with firm slack knowledge stock. I control for firm slack knowledge base as a time-varying stock of patents (aggregated by application date) held by each firm in each component technology lagged by 1 year.

Firm ownership status

Prior research leads us to expect research effort to vary systematically by ownership status of firms. We expect private firms (as compared with public firms) to be smaller and have smaller R&D budgets that will limit their ability to patent their inventions. Also firms filing fewer patents will be unable to exploit economies of scale in patent attorney fees and other administrative overheads. Similarly, foreign firms are likely to have different propensities to patent. Only the larger and more multinational of foreign firms are likely to patent in the United States. Consequently, we measure ownership status as two time-invariant variables: *private*, taking a value of 1 if the patenting firm is private; and *public*, taking a value of 1 if the firm is publicly traded in U.S. capital markets. Foreign firms are the omitted category.

DOI: 10.1002/smj

Time

Prior research suggests that patenting propensities have varied over time (Kortum, 1993). This may be due to an altered regulatory environment, changing appropriability conditions, and so on. I control for patent application year using time dummies or time trend.

Analysis

The dependent variable employed in this study is a proportion, and if the inventive effort allocated to all five components is included in the analysis the total inventive effort for a firm year will sum to one. This property of the dependent variable creates two estimation challenges. First, when the dependent variable is a proportion, estimation using ordinary least squares (OLS) results in the predicted values falling outside the range of the dependent variable. This poses problems in estimating elasticities of the coefficients. This problem, however, is relatively easily fixed. One solution is to logit-transform the proportions and estimate the model via OLS. A second solution is to estimate a two-limit tobit specifying censoring of the dependent variable at 0 and 1, respectively. I implement the former as the primary analyses in the paper and report the latter in the robustness section.

The second problem, i.e., the dependent variable in a firm-year summing to one, causes the errors within panels to be correlated. This problem, which is harder to fix, is commonly encountered in political science in studies predicting relative vote shares in multiparty elections. Recently, Katz and King (1999) developed what they claimed to be the first solution to this estimation challenge. Their approach to deal with the nonindependence of proportions (i.e., increase in the proportion of inventive effort allocated to one component comes at the expense of effort allocated to one or more other components) involves transforming the dependent variable into log odds-ratios and maximizing a likelihood function over a multivariate t-distribution. This approach cannot easily be adapted in this study simply because, as the authors themselves note, it does not scale up easily to models with greater than three groups without resorting to numerical approximations (Katz and King, 1999: 25, 30). I sought to deal with this difficulty in two ways. First, I included data for the R&D effort in only four of the five component categories (system board was the omitted category). This somewhat mitigates the problem since the proportions in a firm year will not sum to zero. From the standpoint of the study, it was also convenient since the system board was never identified as a constraint during the 1981–98 time period. Second, I estimated the equation using generalized least squares (GLS) allowing for errors to be correlated within panels (Davidson and MacKinnon, 1993). While this does not solve all the problems, it appears to be the most feasible solution given the current state of the art. Thus, the basic model specification was the following:

$$\ln\left(\frac{Y_{itj}}{1-Y_{itj}}\right) = \alpha_t + \gamma_j + C_{j(t-1)} + \mathrm{FS}_i \times C_{j(t-1)} + \beta_1 FS_i \times C_{j(t-1)} \times \mathrm{KS}_{i(t-1)} + \beta_{2n} X'_{i(t-1)} + \beta_{3n} X''_{j(t-1)} + \varepsilon_{itj}$$

where Y_{itj} is the allocation of inventive effort by firm i, in year t, in component class j; α_t are the year dummies; γ_j are the component dummies; C captures the constraint; FS is the slack firm dummy, KS is the knowledge stock in the slack components, $X'_{i(t-1)}$ is the vector of firm-specific covariates lagged 1 year; and, $X''_{j(t-1)}$ is the vector of component-specific covariates lagged 1 year. In all models I report robust standard errors, adjusted for multiple observations per firm year.

RESULTS

Table 2 reports the descriptive statistics and correlations for the variables in the analyses. The dependent variable, inventive effort, is the logit-transform of the proportion of patents in each component product, which is why it has a negative mean. Similarly, all the proportion variables that are logged have negative values. Examining the correlation matrix suggests that all the correlations are in the expected direction. Number of employees, knowledge base, and sales are highly correlated ($\gamma > 0.65$). This creates some prima facie concern about possible multicollinearity in estimation. Note, however, that the correlation matrix does not correct for serial correlation among the variables, leading to somewhat inflated coefficients. The estimation procedures employed in the regression models attempt to control for serial correlation.

Table 2. Descriptive statistics and correlation matrix of variables employed in estimation (N = 4639)

	Mean	S.D.	Min.	Max.	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
(1) Inventive effort	−3.54	4.17	−6.91	6.91	1.00									
(2) Constraint	0.16	0.37	0.00	1.00	—	1.00								
(3) Constraint × Slack	0.07	0.25	0.00	1.00	0.08	0.62	1.00							
(4) Constraint × Slack × Firm knowledge stock	−0.07	1.10	−4.61	7.56	0.16	−0.15	−0.25	1.00						
(5) R&D intensity	5.78	2.90	0.38	10.47	0.29	0.12	0.08	−0.01	1.00					
(6) ln(Sales)	9.67	1.51	5.15	12.31	0.30	0.11	0.07	0.00	0.74	1.00				
(7) ln(Employees)	11.83	0.65	10.33	12.76	0.34	0.03	0.06	0.04	0.65	0.68	1.00			
(8) ln(Knowledge stock)	9.44	0.72	7.34	11.05	0.20	−0.08	−0.02	0.05	0.12	0.54	0.62	1.00		
(9) ln(Firm knowledge stock)	−1.14	4.19	−4.61	8.55	0.63	−0.03	0.00	0.24	0.23	0.24	0.25	0.17	1.00	
(10) Firm founding year	1974.31	15.92	1914.00	1996.00	−0.07	0.02	0.03	−0.07	0.02	0.13	0.04	0.16	−0.31	1.00

Correlations greater than 0.025 are significant at the 5% level.

DOI: 10.1002/smj

Table 3. Logit-transformed OLS regression estimates

Model	1	2	3	4
	Inventive effort	Inventive effort	Inventive effort	Inventive effort
Constraint		0.339*	−0.450*	−0.460**
		(0.143)	(0.174)	(0.175)
Constraint × Slack			2.009**	2.133**
			(0.290)	(0.298)
Constraint × Slack × Firm knowledge stock				0.106*
				(0.052)
ln(Firm knowledge stock)	0.635**	0.635**	0.628**	0.619**
	(0.049)	(0.049)	(0.048)	(0.045)
R&D intensity	0.004	0.013	0.014	0.017
	(0.081)	(0.081)	(0.080)	(0.080)
ln(Employees)	−0.744+	−0.806*	−0.944*	−0.973*
	(0.400)	(0.403)	(0.396)	(0.393)
ln(Knowledge stock)	2.002**	2.080**	2.167**	2.185**
	(0.341)	(0.344)	(0.336)	(0.336)
ln(Sales)	0.019	0.026	0.091	0.096
	(0.227)	(0.228)	(0.221)	(0.218)
Concentration	−1.073	−1.073	−0.844	−0.826
	(0.818)	(0.818)	(0.818)	(0.806)
Ownership dummies	Incl.	Incl.	Incl.	Incl.
Component dummies	Incl.	Incl.	Incl.	Incl.
Year dummies	Incl.	Incl.	Incl.	Incl.
Constant	−10.361**	−10.442**	−10.350**	−10.237**
	(2.290)	(2.288)	(2.229)	(2.230)
Observations	4639	4639	4639	4639
F-value	22.52**	21.80**	23.33**	24.63**
R^2	0.408	0.409	0.417	0.418

Robust standard errors in parentheses.
+ Significant at 10%; * significant at 5%; ** significant at 1%.

Table 3 reports the results of a series of regression models. For ease of exposition, I report and interpret only the logit-transformed OLS as the main results. The robustness results with alternative specifications are reported and discussed in the next subsection to provide confidence in the main results reported in Table 3 (see supplementary analyses provided in the Appendix). All models are estimated on a sample of inventive effort in four components (system board is excluded), include year dummies, and report robust standard errors corrected for multiple observations per firm-year. Model 1 estimates an OLS model with only the control variables. The overall model is highly significant and the sign and significance of the coefficients on the control variables are all as expected.[6] I expected significant differences in inventive effort allocation across public, private, and foreign firms. The expectation was largely borne out. The coefficients for the ownership dummies were positive and significant. Similarly, the component industry dummies were also jointly significant, suggesting that the allocation of inventive effort is heterogeneous across component industries. This is not surprising given that there are likely to be significant differences in appropriability and R&D incentives across the component industries.

Model 2 includes the constraint component variable. The coefficient on the constraint component dummy is positive and statistically significant. Since the dependent variable was logit-transformed, interpreting the coefficient on the constraint component involves transforming the coefficient back to its original scale. Generating the predicted values of the logit-transformed dependent variable for the two values of the constraint component dummy and transformed back

[6] Testing for multicollinearity, I found that the highest VIF (variance inflation factor) was 2.23 and the mean for all variables was 1.6, suggesting that multicollinearity was not a serious issue (Chatterjee, Hadi, and Price, 2000).

to the original scale, I found that components that are constraints to system performance on average receive about 13 percent higher inventive effort as compared with slack components. This provides strong support for Hypothesis 1 that component constraints act as triggers for R&D investment.

Model 3 adds the slack firm times constraint component dummy to test Hypothesis 2. Note that the coefficient on the constraint dummy turns negative. Interpreting this coefficient by itself is misleading because of the interaction effect. Accurate interpretation involves computing the joint impact of the main effect and the interaction effect. Once again transforming the coefficients on the constraint dummy and the interaction dummy suggests that much of the investment in resolving component constraints comes from the slack component firms. On average, the slack component firms shift about 8.5 percent of their R&D effort toward resolving component constraints. This provides strong support for Hypothesis 2.

Finally, Model 4 adds the interaction of the slack firm dummy, constraint component dummy, and the firm knowledge stock dummy. The coefficient on the interaction term is positive and significant. Transforming the coefficients to their original scales, I find that a 1 percent increase in firm knowledge stock in slack components increases the investment of slack firms in the constraint component by 0.8 percent. This provides strong support for Hypothesis 3.

Robustness tests

In light of the specification and estimation concerns outlined in the Analysis section, I ran a series of alternative models that are presented in the Appendix. Model A presents the results of an OLS model with the raw proportions as the dependent variable. Model B presents the results of the two-limit tobit. The signs and significance of the coefficients are in accord with the main results presented in Table 3. All three hypotheses continue to be supported.

The models discussed thus far were estimated on a sample that excluded the fifth component (system board) to somewhat mitigate the problem of contemporaneous correlation in the dependent variable for a firm year. Since this fix does not completely eliminate the problem, I sought to estimate the main model in Table 3 using GLS and allowing for panel-specific heteroskedasticity. Model C in the Appendix presents these results. The coefficient estimates for all the variables are generally consistent with the estimates from the previous models. Transforming the estimates back to the original scale as in the case of Model 2 suggests that the component that is a constraint to system performance receives 9.6 percent greater inventive effort than slack components. The results taken together provide convergent evidence that firm-level inventive effort does indeed respond to component constraints in complex product systems.

All of the models reported thus far assume that the patenting propensity of firms is uniform across component industries. In other words, when a firm does not allocate inventive effort to a component industry in a given year, the dependent variable for that firm–component–year will be zero. The models above assume that the process that is driving the nonallocation of inventive effort (i.e., zero effort) is no different from the process driving the magnitude of inventive effort (i.e., effort greater than zero). However, this is unlikely to be the case since patenting propensities vary systematically both by firm (Hausman *et al*., 1984), and industry (Cohen *et al*., 2000). In other words, the decision to allocate R&D effort in a given component industry may be endogenous to the firm and ignoring this endogeneity can pose an omitted variable problem.

To control for this, I estimated a two-step Heckman model. I first estimated, using a binary probit model, the likelihood that a firm will allocate R&D effort in a particular component technology. The dependent variable is coded 1 if it allocates inventive effort in a component industry in a given year and 0 otherwise. This produced an estimate of a nonselection hazard for each observation in the sample. In the second step, I estimated an OLS model including the nonselection hazard. The results of this estimation are presented in Models D and E. Model D presents the selection model that estimates the likelihood that a firm will allocate R&D effort in a component technology. As can be seen, the nonselection hazard (lambda) is statistically significant, suggesting selection bias in the data. Model E presents the estimation model after correcting for the selection bias. The results are largely consistent with the results reported in the previous models. This provides some confidence that the effect of constraint-induced innovation is present even when I control for endogeneity in the allocation of R&D effort to a component industry.

DOI: 10.1002/smj

Because I confined this study to only component innovation and excluded architectural innovation, the possibility of an important alternative explanation is raised. If architectural change is the primary impetus for component innovation and the timing of architectural change coincides with the timing of the identification of constraints, I would be wrongly attributing my results to the constraints explanation, whereas architectural change may be the true underlying reason. I sought to deal with this potential omitted variable empirically. The personal computer architecture is largely defined by the width of the data bus (Messmer, 1997). The data bus architecture has changed three times during the timeframe of this study: 1983 (change from 8-bit to 16-bit), 1985 (change from 16-bit to 32-bit), and 1993 (change from 32-bit to 64-bit). I dropped all identification of constraints in the years 1983, 1985, and 1993 (see Table 1 for years in which component constraints were identified) and reran all the models in Table 3. The results were unchanged, providing confidence that coincidental architectural change is unlikely to be driving the results.

In addition to the models reported in the paper, I performed a variety of other robustness checks. The models above do not account for unobserved, but systematic firm differences in patenting propensities. Inclusion of the firm-specific effects in the above models would preclude the inclusion of the S-firm, component industry, and firm ownership dummies, respectively. For additional robustness, I estimated fixed-effects and random-effects versions of Model 2. The results were largely similar. The component constraint variable continues to be positive and significant, though the magnitude declines marginally to 7.8 percent. Overall the results seem robust to these alternative model specifications.

A second concern was the presence of autocorrelation in the panel data. I plotted the residuals from Model 4 in Table 3 for year t against year $t - 1$. The plot was indicative of a marginally positive first-order serial correlation in the data. To examine the significance of serial correlation on the reported results, I extracted a balanced sample of firms and re-estimated Model 4 in the paper allowing for first-order, within-panel autocorrelation. The estimates were largely identical to that of Model 4, providing further confidence that serial correlation is unlikely to be driving the results.

Finally, there is some concern that firms with large R&D budgets such as Intel Corp. may be driving the results. I reran Model 4 eliminating such prominent firms and the results were qualitatively similar, providing confidence that the results are not driven by such outliers.

DISCUSSION

The study reported here makes some useful theoretical and empirical contributions to the literature on firm R&D scope. The primary contribution of the paper is in drawing out the relationship between technological characteristics and firm incentives to invest in innovation. The paper argues and empirically demonstrates that interactions between components in a product system conditions the R&D incentives of firms and also that the incentives are increasing in their prior investments or capabilities. The empirical study provided strong support for all three hypotheses advanced in the paper. In summary, I would like to emphasize three principal contributions of this study.

First, the empirical results confirm the counterintuitive prediction that slack component firms have a strong incentive to resolve component constraints. Furthermore, the incentive of slack component firms to invest in the constraint component is increasing in their sunk R&D investments in slack components. This highlights an apparent paradox of modularity. Rather than increase specialization and reduce the scope of firm R&D activities, modularity creates an impetus for firms in a product system, regardless of the component product market in which they compete, to resolve component constraints and thus expand their R&D scope beyond their component product markets. At a more general level, the apparent paradox highlighted here is hardly surprising. By definition, what makes large-scale systems complex is the nonlinearities inherent in their behavior (Cohen and Axelrod, 1999). By this token, it is not surprising that modularity, as with any organizing principle in complex systems, comes with its own trade-offs. On the one hand, modularity accelerates innovation by promoting specialization and facilitating autonomous research efforts (Ethiraj and Levinthal, 2004b). But on the other hand,

DOI: 10.1002/smj

it increases the interdependence of firm component R&D efforts; realizing returns from component R&D is often contingent on concomitant developments in other components that are beyond the control of individual firms. This, in turn, creates the impetus for firms to expand R&D efforts beyond the component product markets in which they operate. This intuition also extends in important ways the research on why firm R&D scope is often much broader than their product scope (Brusoni *et al*., 2001). This paper broadens the explanation for breadth of R&D scope to all firms in complex product systems rather than just systems integrators.

Second, the empirical results suggest that constraint component(s) receive about 13 percent higher inventive effort on average, as compared with slack components. I also found that a significant proportion of inventive effort is stable (about 40%), i.e., not induced by technical imbalances. This leads to the question of whether constraint-induced innovation is trivial enough to be ignored by R&D managers. The coefficient on the constraint component is net of the variance in allocation of inventive effort explained by R&D expenditure, scale, age, and so on. It may be explaining unobservables that constitute inventive effort (e.g., managerial attention, management commitment). It is hard to attach economic value to such factors, making it hard to quantify the economic impact of constraint-induced technical change.

However, some qualitative assessment is possible. The proportion of patents in each component technology is the proxy for allocation of inventive effort. Therefore, a 10 percent shift (net of R&D, scale, etc.) in allocation in a particular year means that the patents applied for in the constraint component domain increases by 10 percent. For instance, assume a system with two components, A and B, with a total of 100 patents. Component A accounts for 20 patents (B accounts for the remaining 80) in period 1 when it is identified as a constraint. In period 2, a shift in allocation means that component A will witness an eight-patent increase (10% of 80 patents in component B), signifying a year-over-year increase from 20 to 28 patents. Conversely, if component B is a binding constraint, it will witness a year-over-year increase of two patents (10% of 20 patents in component A). This suggests that components that have a large accumulated knowledge base are less likely to be radically affected by the emergence of constraints. In contrast, components with a relatively small knowledge base can witness large technological changes that can, in turn, significantly affect the trajectories of product systems. Under such circumstances, ignoring constraint-induced innovation can prove detrimental to R&D managers. In other words, my study suggests that the emergence of constraints in relatively less important and innocuous components might trigger radical changes in them that in turn present implications for the more important components as well.

Lastly, the paper has taken the first step in directly testing the constraint-induced innovation hypothesis. The induced innovation hypothesis is mostly supported by historical accounts of technological change and innovation (Rosenberg, 1969, 1974). Prior large-sample empirical work in this area is concentrated in the agriculture sector, where researchers have tried to examine factor substitution biases rather than induced innovation. R&D activity in complex product systems provides an ideal context for testing the induced innovation hypothesis. I examined the changes in relative R&D allocation to different components in response to the emergence of 'technical imbalances' caused by component constraints. This enabled the first large-sample empirical test of Rosenberg's (1969) assertion that technical imbalances or constraints induce shifts in the allocation of inventive effort. A test of this idea is important to the extent that implications of ignoring constraints might threaten the survival of incumbent firms. It appears that the security of specialization that modularity affords does not extend to R&D investment choices. It is important for firms to recognize both the power and limits of modular architectures in functionally interdependent product systems.

Lastly, there is the important question of generalizability of the arguments and findings in the paper. In this respect, Pil and Cohen (2006) suggest that an important distinction in modular systems is between open and closed architectures. Whereas open systems have common interface specifications (such as the PC), closed systems do not (e.g., automobiles and bicycles). This raises the question of whether my findings generalize only to open, modular systems such as PCs. There are two necessary conditions for my results to hold: (1) modularity in design of components (as distinct from modularity in production and modularity in

DOI: 10.1002/smj

use); and (2) competition in the component markets (i.e., entry and exit). These two conditions are necessary for the principal mechanisms in my hypotheses to become salient, i.e., functional interdependencies between components and imbalances in their rates of technical progress. If these two conditions are met, I expect my results to carry to other modular products regardless of whether they are characterized by open or closed architectures.

In conclusion, the paper has documented interesting patterns of innovation and change in complex product systems. However, the paper examined only one small piece of the larger puzzle of the coordination of economic activity in complex systems. A subject worthy of future research is a comprehensive examination of the trade-offs posed by modular design principles. Examining how modularity affects other important organizational choices and processes will aid in a fuller understanding of the managerial challenges in dealing effectively with complexity (Ethiraj and Levinthal, 2004a).

ACKNOWLEDGEMENTS

Financial support from the Huntsman Center for Global Competition and Innovation at the Wharton School is gratefully acknowledged. I thank two anonymous reviewers, Gautam Ahuja, George Day, Dan Levinthal, Hart Posen, Harbir Singh, Jitendra Singh, Sidney Winter, and seminar participants at the Wharton School, Boston University, University of Minnesota, University of Southern California, Harvard Business School, MIT Sloan School of Management, University of Michigan, London Business School, Carnegie Mellon University, and New York University for useful comments. I also thank executives at IBM, Intel, Texas Instruments, and the staff at CHI Research and the Office of Technology and Forecasting for help with mapping patent classes to component industries.

REFERENCES

Almeida P, Kogut B. 1999. Localization of knowledge and the mobility of engineers in regional networks. *Management Science* **45**(7): 905–917.

AMD. 2006. AMD 3-year technology outlook. http://www.amdcompare.com/techoutlook/ [26 September 2006].

Arrow K. 1962. Economic welfare and the allocation of resources for invention. In *The Rate and Direction of Inventive Activity*, Nelson R (ed). Princeton University Press: Princeton, NJ; 609–625.

Baldwin CY, Clark KB. 2000. *Design Rules: The Power of Modularity*. MIT Press: Cambridge, MA.

Brusoni S, Prencipe A, Pavitt K. 2001. Knowledge specialization, organizational coupling, and the boundaries of the firm: why do firms know more than they make? *Administrative Science Quarterly* **46**(4): 597–621.

Cadambi B. 2006. PCI-Express Case study. Intel Corp. http://developer.intel.com/standards/case/case_pci.htm [8 May 2006].

Chatterjee S, Hadi AS, Price B. 2000. *Regression Analysis by Example* (3rd edn). Wiley: New York.

Chen A, Johnson J, Suhrstedt A. 2002. Introduction to AGP-8X. Advanced Micro Devices: http://www.amd.com/us-en/assets/content_type/white_papers_and_tech_docs/AGP8X_26292A_PM_revision.pdf [accessed 26 September 2006].

Christensen C. 1997. *The Innovator's Dilemma: When New Technologies Cause Great Firms to Fail*. Harvard Business School Press: Boston, MA.

Clark KB, Fujimoto T. 1991. *Product Development Performance: Strategy, Organization, and Management in the World Auto Industry*. Harvard Business School Press: Boston, MA.

Cohen M, Axelrod R. 1999. *Harnessing Complexity: Organizational Implications of a Scientific Frontier*. Free Press: New York.

Cohen WM, Levinthal DA. 1990. Absorptive capacity: a new perspective on learning and innovation. *Administrative Science Quarterly* **35**(1): 128–152.

Cohen WM, Nelson RR, Walsh J. 2000. Protecting their intellectual assets: appropriability conditions and why U.S. manufacturing firms patent (or not). Working paper, National Bureau of Economic Research, Boston, MA.

Dasgupta P. 1988. Patents, priority and imitation or, the economics of races and waiting games. *Economic Journal* **98**(389): 66–80.

Davidson R, MacKinnon JG. 1993. *Estimation and Inference in Econometrics*. Oxford University Press: Oxford, U.K.

Dierickx I, Cool K. 1989. Asset stock accumulation and sustainability of competitive advantage. *Management Science* **35**(12): 1504–1512.

Ethiraj SK, Levinthal DA. 2004a. Bounded rationality and the search for organizational architecture: an evolutionary perspective on the design of organizations and their evolvability. *Administrative Science Quarterly* **49**(3): 404–437.

Ethiraj SK, Levinthal DA. 2004b. Modularity and innovation in complex systems. *Management Science* **50**(2): 159–173.

Garud R, Kumaraswamy A. 1993. Changing competitive dynamics in network industries: an exploration of Sun Microsystems' open systems strategy. *Strategic Management Journal* **14**(5): 351–369.

Garud R, Kumaraswamy A. 1995. Technological and organizational designs for realizing economies of substitution. *Strategic Management Journal*, Summer Special Issue **16**: 93–109.

Griliches Z (ed). 1984. *R&D, Patents, and Productivity*. University of Chicago Press: Chicago, IL.

Hausman J, Hall B, Griliches Z. 1984. Econometric models for count data with an application to the patents–R&D relationship. *Econometrica* **52**(4): 909–938.

Hoetker G. 2006. Do modular products lead to modular organizations? *Strategic Management Journal* **27**(6): 501–518.

Intel Corp. 2006. Intel PCI and PCI Express: case study. http://developer.intel.com/standards/case/case_pci.htm [8 May 2006].

Katz JN, King G. 1999. A statistical model for multiparty electoral data. *American Political Science Review* **93**(1): 15–32.

Khanna T. 1995. Racing behavior: technological evolution in the high-end computer industry. *Research Policy* **24**(6): 933–959.

Kortum S. 1993. Equilibrium R&D and the patent–R&D ratio: U.S. evidence. *American Economic Review* **83**(2): 450–457.

Langlois R, Steinmueller WE. 1999. The evolution of competitive advantage in the worldwide Semiconductor industry, 1947–1996. In *The Sources of Industrial Leadership*, DC Mowery RR Nelson (eds). Cambridge University Press: New York; 19–78.

Levin RC, Klevorick AK, Nelson RR, Winter SG. 1987. Appropriating the returns from industrial research and development. *Brookings Papers on Economic Activity* **3**: 783–820.

Malerba F, Nelson R, Orsenigo L, Winter S. 1999. History-friendly models of industry evolution: the computer industry. *Industrial and Corporate Change* **8**(1): 3–40.

Messmer H-P. 1997. *The Indispensable PC Hardware Book*. Addison-Wesley: Reading, MA.

Nelson RR. 1959. The economics of invention: a survey of the literature. *Journal of Business* **32**(2): 101–127.

Nelson RR. 1961. Uncertainty, learning, and the economics of parallel research and development efforts. *Review of Economics and Statistics* **43**(4): 351–364.

Nelson RR, Winter SG. 1982. *An Evolutionary Theory of Economic Change*. Belknap Press of Harvard University Press: Cambridge, MA.

Patel P, Pavitt K. 1997. The technological competencies of the world's largest firms: complex and path-dependent, but not much variety. *Research Policy* **26**(2): 141–156.

Pavitt K, Robson M, Townsend J. 1989. Technological accumulation, diversification, and organization in U.K. companies, 1945–1983. *Management Science* **35**(1): 81–99.

Pil FK, Cohen SK. 2006. Modularity: implications for imitation, innovation, and sustained advantage. *Academy of Management Review* **31**(4): 995–1011.

Prencipe A. 1997. Technological capabilities and product evolutionary dynamics: a case study from the aero engine industry. *Research Policy* **25**(8): 1261–1276.

Rosenberg N. 1969. The direction of technological change: inducement mechanisms and focusing devices. *Economic Development and Cultural Change* **18**(1): 1–24.

Rosenberg N. 1974. Science, invention, and economic growth. *Economic Journal* **84**(333): 90–108.

Rumelt RP. 1986. *Strategy, Structure, and Economic Performance* (rev. edn). Harvard Business School Press: Boston, MA.

Sahal D. 1981. *Patterns of Technological Innovation*. Addison-Wesley: Reading, MA.

Sanchez R, Mahoney JT. 1996. Modularity, flexibility, and knowledge management in product and organization design. *Strategic Management Journal*, Winter Special Issue **17**: 63–76.

Schilling M. 2000. Toward a general modular systems theory and its application to interfirm product modularity. *Academy of Management Review* **25**(2): 312–334.

Schilling M, Steensma KH. 2001. The use of modular organizational forms: an industry-level analysis. *Academy of Management Journal* **44**(6): 1149–1168.

Schmookler J. 1966. *Invention and Economic Growth*. Harvard University Press: Cambridge, MA.

Schumpeter JA. 1950. *Capitalism, Socialism, and Democracy*. Harper: New York.

Shibata T, Yano M, Kodama F. 2005. Empirical analysis of evolution of product architecture Fanuc numerical controllers from 1962 to 1997. *Research Policy* **34**(1): 13–31.

Simon HA. 1962. The architecture of complexity. *Proceedings of the American Philosophical Society* **106**(6): 467–482.

Sorenson J, Stuart T. 2000. Aging, obsolescence, and organizational innovation. *Administrative Science Quarterly* **45**(1): 81–112.

Trajtenberg M. 1990. A penny for your quotes: patent citations and the value of innovations. *RAND Journal of Economics* **21**(1): 172–187.

Ulrich KT. 1995. The role of product architecture in the manufacturing firm. *Research Policy* **24**(3): 419–440.

Ulrich KT, Eppinger SD. 1999. *Product Design and Development* (2nd edn). McGraw-Hill: New York.

Winter SG. 1984. Schumpeterian competition in alternative technological regimes. *Journal of Economic Behavior and Organization* **5**(3–4): 287–320.

DOI: 10.1002/smj

APPENDIX: ROBUSTNESS TESTS

Model	A OLS	B Two-limit Tobit	C Logit transformed GLS	D Heckman—selection	E Heckman—Main
Constraint	−0.041**	−0.104**	−0.144		−0.610*
	(0.014)	(0.035)	(0.136)		(0.241)
Constraint × Slack	0.198**	0.386**	1.184**		2.728**
	(0.025)	(0.045)	(0.191)		(0.340)
Constraint × Slack × Firm knowledge stock	0.010*	0.004	0.082*		0.169*
	(0.005)	(0.008)	(0.034)		(0.068)
ln(Firm knowledge stock)	0.036**	0.101**	0.555**	0.701**	0.361**
	(0.004)	(0.003)	(0.010)	(0.036)	(0.034)
R&D intensity	−0.002	0.003	0.125**		−1.056*
	(0.007)	(0.010)	(0.041)		(0.506)
ln(Employees)	−0.093**	−0.170**	−0.792**		0.362
	(0.034)	(0.051)	(0.217)		(0.366)
ln(Knowledge stock)	0.228**	0.402**	1.802**		1.656**
	(0.029)	(0.053)	(0.266)		(0.375)
ln(Sales)	0.025	0.018	0.265**		−0.030
	(0.021)	(0.024)	(0.097)		(0.172)
Concentration	−0.075	0.008	−0.080	−0.872+	
	(0.068)	(0.077)	(0.407)	(0.482)	
Firm founding year				0.024**	
				(0.004)	
Non-selection hazard				−0.544**	
				(0.209)	
Ownership dummies	Incl.	Incl.	Incl.	Incl.	N.A.
Component dummies	Incl.	Incl.	Incl.	Incl.	N.A.
Year dummies	Incl.	Incl.	Incl.	Incl.	N.A.
Constant	−1.045**	−1.583**	−11.932**	−45.880**	−17.184**
	(0.213)	(0.287)	(1.209)	(8.832)	(1.606)
Observations	4639	4639	4639	4639	4639
R^2	0.337				

Robust standard errors in parentheses.
+ Significant at 10%; * significant at 5%; ** significant at 1%.

DOI: 10.1002/smj