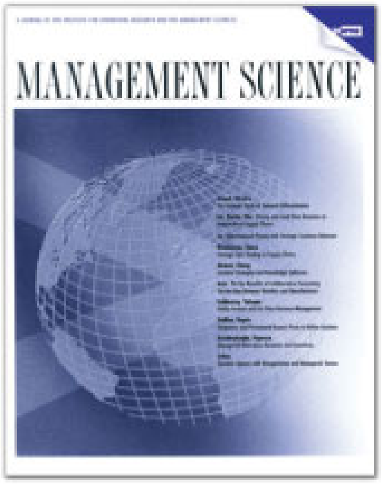

Management Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Local R&D Strategies and Multilocation Firms: The Role of Internal Linkages

Juan Alcácer, Minyuan Zhao,

MANAGEMENT SCIENCE
Vol. 58, No. 4, April 2012, pp. 734–753
ISSN 0025-1909 (print) | ISSN 1526-5501 (online)

informs
http://dx.doi.org/10.1287/mnsc.1110.1451

Local R&D Strategies and Multilocation Firms: The Role of Internal Linkages

Juan Alcácer
Harvard Business School, Harvard University, Boston, Massachusetts 02163,
jalcacer@hbs.edu

Minyuan Zhao
Ross School of Business, University of Michigan, Ann Arbor, Michigan 48109,
myzhao@umich.edu

This study looks at the role of internal linkages in highly competitive clusters. We argue that, in addition to serving as a mechanism for sourcing knowledge, strong internal linkages help firms increase internalization and create higher levels of technological interdependence across firm locations. Firms with strong networks of internal linkages are able to maintain tighter control over local innovation and reduce the risk that knowledge outflows will advantage competitors in clusters. Our empirical analysis of the global semiconductor industry shows that industry leaders intensify internal linkages across locations when they collocate with direct market competitors, but not when they collocate with innovators in the same technological field. We also find that internal linkages are associated with more knowledge flow within firms and less knowledge expropriation by collocated competitors. Our results suggest that future research in cluster innovation should consider the critical role of multilocation firms, their internal organization across clusters, and their responses to technological and market competition in clusters.

Key words: technology clusters; knowledge spillover; internalization; appropriability
History: Received August 12, 2009; accepted July 25, 2011, by Bruno Cassiman, business strategy. Published online in *Articles in Advance* December 2, 2011.

1. Introduction

Marshall (1920) suggests that clusters reduce costs for collocated firms by providing convenient access to skilled labor, specialized suppliers, and knowledge spillovers. More recently, clusters have been credited with facilitating innovation in high-technology industries (Saxenian 1994) and creating competitive advantages for participating firms. Large numbers of firms and research institutions engaged in innovative activities in close proximity will facilitate knowledge flow by increasing opportunities for interpersonal interactions (Audretsch and Feldman 1996) and labor mobility (Almeida and Kogut 1999), thereby providing a fertile environment for knowledge exchange.

Of course, locations that attract large numbers of firms will also create more competition for critical resources and knowledge. Porter (1998, p. 227) emphasizes the "vigorous competition among locally based rivals." Shaver and Flyer (2000) point out that collocation is not universally beneficial because leading firms might lose more to proximate competitors than they gain. Knowledge outflows that create advantages for competitors can weaken the competitive edge of leading innovators and compromise their ability to appropriate value from research and development (R&D).

This paper continues the inquiry into innovation in clusters by examining a critical paradox: Joining clusters is more perilous for leading firms, and yet leading firms in high-technology industries flock to clusters. What allows these leaders to benefit from clusters without losing their technological edge? We address this question by studying how leading firms operating in multiple clusters leverage internal linkages *across* clusters to improve knowledge appropriation. Stronger internal linkages across clusters can enhance appropriation by simultaneously *increasing* firms' ability to internalize knowledge and *decreasing* the risk that knowledge outflows will advantage competitors. Thus, firms with strong networks of internal linkages can leverage the resource and knowledge benefits of competitive local environments while *also* appropriating value from firm innovations.

As we show in §2.1, previous research about the benefits and risks of collocating generally overlooks the impact of geographically dispersed organizations. We point out that many of the dominant players in clusters are large, multilocation firms known for their ability to mobilize and integrate knowledge on a global basis. It follows that a firm's innovation strategy in one cluster may be influenced by its internal linkages with other firm locations. How firms organize their innovation activities internally and across locations will ultimately affect their ability to appropriate value from innovations in any given location.

We review appropriation strategies in the strategy and innovation literatures in §2.2, paying particular attention to the strategies firms use to retain control of proprietary knowledge, a process we call *knowledge appropriation*. Most of these mechanisms hinge on reducing knowledge outflows to competitors or reducing competitors' incentives to imitate. However, by examining firms in the abstract, these literatures tacitly assume that firms develop knowledge appropriation mechanisms independent of their locations. We point out that, because knowledge spillovers are more likely when multiple firms share a geographic space, firms operating in competitive local environments should craft knowledge appropriation strategies with their locations in mind.

We propose in §2.3 that internal linkages improve knowledge appropriation in highly competitive clusters. Specifically, we suggest that a firm innovating in multiple locations can enhance knowledge appropriation by using internal linkages to (1) closely monitor R&D activities across firm locations and (2) increase interdependence between locations. Our arguments complement the common view in the literature that internal linkages are solely effective for absorbing and integrating external knowledge, a process we refer to as *knowledge sourcing*.

Internal linkages take many forms. Following qualitative interviews with managers in the semiconductor industry, we identified one type of internal linkage that appears critical to firms' knowledge appropriation strategies, and that can be tested effectively with our data set of multilocation semiconductor firms. We name these linkages *cross-cluster teams* and propose that they will (1) be employed more often by firms operating in clusters where knowledge appropriation is at risk, (2) enhance the flow of proprietary knowledge between firm locations, and (3) reduce the value of any proprietary knowledge that collocated competitors obtain.

Our empirical setting is innovation in the global semiconductor industry from 1998 to 2001, described in detail in §3.1. Specifically, we analyze three innovation traits—the existence of cross-cluster teams, the number of cross-cluster self-citations in patents, and the number of citations by local competitors—exhibited by the top 16 innovators in 25 clusters located worldwide. The semiconductor industry is an ideal setting for our research because innovation is traceable by patents, the technology is used in multiple product markets, and most firms innovate across locations. We focus on industry leaders operating in multiple locations because (1) they are most likely to produce cutting-edge technologies, (2) they have a high incentive to protect proprietary knowledge, and (3) operating in multiple locations makes internal linkages a feasible mechanism for enhancing appropriation. We depart from previous research that defines clusters in terms of employment density in predetermined geographic units. Instead, we define the contours of individual clusters using a mathematical algorithm whose input is the location of a patent's inventor(s). This approach, elaborated in §3.2, allows us to identify locations that are rich in innovation rather than rich in production. We call these innovation-rich clusters *technology clusters*. The remainder of our methods are described in §3.3 (models), §3.4 (dependent variables), and §3.5 (independent variables).

Section 4 presents evidence supporting our proposition that firms leverage internal linkages for knowledge sourcing *and* appropriation, with much stronger results for the latter. Specifically, leading firms in our sample employed cross-cluster teams more often when located in clusters with a large number of market competitors, even after controlling for the opportunity of local knowledge sourcing. We also find evidence that leveraging internal linkages to enhance appropriation is an effective strategy for firms in clusters. Specifically, technologies produced by cross-cluster teams in our sample were *more likely* to be transferred internally to other locations of the same firm, and *less likely* to underpin future innovation by collocated competitors. Interestingly, whereas the knowledge-sourcing effect is significant among the general population of the cluster, the appropriation effect is significant only with direct market competitors, suggesting that appropriation strategies are highly targeted. This helps to explain how firms can, in fact, pursue knowledge sourcing and knowledge appropriation simultaneously in clusters: knowledge appropriation strategies that target direct market competitors do not prevent a firm from knowledge sourcing with other players in the community. These findings are robust to a battery of checks, which we describe in §5. Finally, §6 concludes the paper with our results, theoretical and empirical contributions, and suggestions for future research.

In sum, our study of knowledge appropriation strategies among multilocation firms in clusters begins to answer the paradox we identified at the outset—how leading firms in high-tech industries can benefit from collocation without risking their technological edge—and makes three important contributions to the strategy, innovation, and international business literatures. First, it advocates for a neglected dimension in the appropriation literature—location—by demonstrating that actual knowledge outflows, and the strategies firms craft to prevent them, are shaped by the locations in which firms operate. We also introduce location to the literature on interactions between firms' technology and product-market strategies, showing that knowledge appropriation and knowledge sourcing are not location free. More specifically, we show that one type of internal linkages, cross-cluster teams, are an important tool for knowledge appropriation in highly competitive clusters and worthy of additional research. We believe this is one of the first papers to analyze the role of internal linkages in knowledge appropriation. Finally, by demonstrating the role of cross-cluster teams in knowledge appropriation strategies, our study suggests that future inquiry into the strategic dynamic of any particular cluster should consider the critical role of multilocation firms and their internal organizations *across* clusters. Empirically, our study contributes a new methodology for defining clusters that is particularly useful for comparing clusters across countries.[1]

2. Theoretical Development

In this section we analyze the features of clusters and the importance of appropriation strategies for firms surrounded by direct competitors. We argue that, among the many appropriation mechanisms discussed in the literature, strong internal linkages enable multilocation firms to integrate local R&D with complementary assets residing elsewhere in the world and thereby strengthen their control over proprietary knowledge gained through innovation.

2.1. Clusters and Firm Heterogeneity

According to Porter (1998), clusters are a prominent feature in the landscape of every advanced economy. Starting with seminal work by Marshall (1920), researches have argued that firms in a cluster benefit from knowledge spillover across organizations, access to specialized labor, and access to specialized intermediate inputs. Among the various activities along the value chain, R&D benefits the most from local knowledge spillover and shows the highest level of geographic concentration (Audretsch and Feldman 1996, Alcácer 2006). Geographic proximity facilitates the transfer of tacit knowledge by enabling frequent interpersonal interactions in social networks (Almeida and Kogut 1999) and local institutions (Gilson 1999).

However, knowledge flows in both directions. Knowledge that flows into the firm (knowledge inflow) may make R&D investment more productive and thus raise the incentive to invest in R&D in clusters. At the same time, knowledge flowing out of the firm (knowledge outflow) may hinder the firm's ability to appropriate value from its own innovations, lowering its incentive to conduct R&D in clusters (Cassiman and Veugelers 2002). Appropriation is particularly at risk when knowledge outflows can be obtained, integrated, and leveraged for advantage by a firm's competitors, an outcome we call *knowledge expropriation*. Industry leaders are particularly at risk from knowledge outflows in clusters because knowledge expropriation can erode a leader's competitive advantage.

Leading firms can move away from clusters to protect their cutting-edge technologies, but this option may not be sustainable or desirable for two reasons. First, even if a leading firm decides to locate apart, it has little control over competitors' subsequent location decisions or the emergence of new firms. If other firms have incentives to cluster around industry leaders, geographic distance offers only temporary protection against knowledge expropriation. Second, if there are crucial resources in the cluster that the firm relies on—such as the talent pool from a local university—moving outside that cluster could seriously compromise the firm's long-term competitiveness (Chung and Alcácer 2002). Hence, preventing collocated competitors from expropriating proprietary knowledge is a strategic consideration leading firms cannot avoid.

One feature that industry leaders *can* control is a geographically dispersed, but closely integrated, innovation network. The literature on clusters traditionally treats all local entities as stand-alone organizations. As a result, interactions among local competitors have been examined without considering firms' extended organizations. Most of the leading firms in high-tech industries are large firms with R&D activities in multiple locations or even multiple countries. As emphasized by Pisano (2006), an industry's appropriation methods are created by the strategic decisions of firms in that industry. Hence, the strategic allocation and integration of innovation activities by multilocation firms will have broad and important implications for how firms interact with other organizations in a cluster.

2.2. Appropriation Mechanisms

The innovation literature discusses a wide range of strategies for knowledge appropriation. These strategies fall into two broad categories: raising barriers to imitation and reducing incentives for imitation.

Firms may raise barriers to imitation by maintaining physical distance from potential imitators (e.g., Shaver and Flyer 2000). They may also implement organizational designs to ensure secrecy and manage access to proprietary information. In fact,

[1] As a result of the work undertaken for this study, the authors will make public location data for patents granted in the United States from 1969 to 2010. The data has been cleaned and contains latitude and longitude for most inventors in the USPTO (United States Patent and Trademark Office) data set.

the 1987 Yale survey and the 1994 Carnegie Mellon survey both identified secrecy as among the most important mechanisms firms use to protect R&D investments (Cohen et al. 2000). Legal devices such as patents and noncompete clauses also effectively increase the costs of imitation and deter information flows. Prior studies have found that employees who are aware of trade-secret handling procedures feel more obligated to protect trade secrets (Hannah 2005). Even the corporate reputation for being "tough" in patent enforcement reduces knowledge outflows associated with employee mobility (Agarwal et al. 2009).

Alternatively, firms can reduce the incentives for imitation. For example, many technologies are valuable only when combined with the right complementary assets, including physical assets, marketing and managerial skills, brand names, know-how, and technological capabilities (Teece 1986, Fosfuri et al. 2008). As a result, R&D is often firm-specific in its intended use, leading to heterogeneity across firms in terms of R&D applications and appropriability of R&D returns by the innovating firms (Helfat 1994). Similarly, when alliance partners compete in the same product or geographic markets, they avoid direct market competition by restricting the scope of their collaboration to R&D activities (Oxley and Sampson 2004). Not surprisingly, the Carnegie Mellon survey found that complementary capabilities and lead time advantages are also important mechanisms for knowledge appropriation (Cohen et al. 2000).

Although these appropriation mechanisms are important in various circumstances, they remain abstract in the sense that they treat firms as operating free of location. Actual knowledge spillovers, and the actions firms take to prevent them, happen in the context of specific locations. In this light, the gap in the knowledge appropriation literature is almost opposite the gap in the cluster literature: the appropriation literature discusses firms' strategic organization without considering location characteristics; the cluster literature emphasizes the role of location but overlooks firms' complex internal organizations. The following discussion explores the interaction between firm location and firm internal organization by focusing on a particular dimension of internal organization: internal linkages across locations.

2.3. Internal Linkages, Knowledge Sourcing, and Knowledge Appropriation

Researchers have long recognized that internal linkages in firms are an effective means of absorbing and integrating external knowledge, or knowledge sourcing. By establishing interactions across divisions or distances, internal linkages facilitate the accumulation and integration of knowledge (Kogut and Zander 1993). Empirical evidence shows that strong internal linkages can improve a firm's knowledge absorption and integration (Gupta and Govindarajan 2000), increase the absorption of external knowledge at dispersed locations (Lahiri 2010), facilitate the transfer of local knowledge back to the parent firm (Frost and Zhou 2005), and improve the overall quality of innovation (Singh 2008).

We argue that, in addition to their knowledge-sourcing effect, internal linkages are a means of knowledge appropriation for multilocation firms. Specifically, we propose two mechanisms through which internal linkages can promote internalization and reduce the threat of knowledge expropriation. First, internal linkages allow firms to maintain effective control over innovative activities at various R&D centers. Second, internal linkages are associated with strong technological interdependence within the firm, making innovations more valuable internally than to outsiders. We explain in detail below.

2.3.1. Internal Linkages as a Mechanism of Control. Firms with strong internal ties can closely monitor the progress of R&D activities at each location to ensure those activities align with firm priorities. Examples include the participation of researchers from other locations in local R&D projects (Nobel and Birkinshaw 1998) or the rotation of managers across units (Edstrom and Galbraith 1977). Through such internal ties, firms gain more comprehensive knowledge of local R&D activities and their outcomes, which allows them to draw clear boundaries for their intellectual properties and to defend them when necessary. In addition, involving team members from other locations discourages the formation of highly localized capabilities. For instance, employee spin-offs with proprietary technologies are much more difficult if the technologies involve researchers from multiple locations. Therefore, strong internal linkages help enhance monitoring and control over local innovation, strengthening knowledge appropriation at the firm level.

Our interviews with R&D and intellectual property (IP) managers from six large multinational firms in the semiconductor industry also suggest that frequent interactions among locations allow valuable innovations to be identified promptly and sometimes transferred to locations with fewer direct competitors, often at a firm's headquarters or primary R&D centers.[2] For instance, an IP manager with a Korean

[2] We conducted unstructured interviews with 11 managers in six major semiconductor firms (IBM, Intel, ST microelectronics, Samsung, Hynix (LG Electronics), and Hitachi) to identify relevant appropriation mechanisms.

firm described the strategic role of internal linkages this way:

> We monitor constantly innovation developed in the subsidiaries. Important projects usually involve personnel from headquarters…(because) they not only coordinate the project, they are also there to check that our subsidiaries follow our IP policy and let us know when we have to file for patents…. Sometimes projects become too important or risky, so we need to bring them home.

2.3.2. Internal Linkages as a Mechanism of Internal Interdependence. Firms with strong internal ties can better integrate innovation—wherever it emerges—with the complementary knowledge and resources within the firm, leading to greater internal interdependence and a stronger competitive position in the product market. Modularity with firm-specific interfaces is one example of internal interdependence. Liebeskind (1996) proposes that firms can isolate various components of the same product so that any project team cannot reproduce the product without help from other teams. Similarly, a multinational firm's local innovations will be less attractive to collocated competitors when those competitors do not have access to critical complementary knowledge. In studies of multinational R&D strategies, Zhao (2006) found that firms with strong internal linkages could safely conduct R&D in environments with weak intellectual property rights protection when their internal organization acted as a substitute for external institutions. As one R&D manager from an American semiconductor manufacturer explained during an interview:

> The majority of our projects involve global collaboration. Some…are about customization, in which case headquarters needs feedback from different locations. Others are large-scale projects with a long chain of components dependent upon each other. A committee comprised of senior technicians, including expats from headquarters, oversees the progress of R&D.

The role of interdependence in knowledge appropriation has also been discussed in more general settings. Using a theoretical model, Rajan and Zingales (2001) explain why flat hierarchies—in which all division managers are required to collaborate with a central unit at the top—are ubiquitous in human-capital-intensive industries such as legal and consulting services. Property rights protection is difficult to enforce when firm resources are intangible. Yet, when firm divisions are highly interdependent, and when the center office controls access to certain key resources, expropriation risks are greatly reduced.

For these reasons, we argue that a firm can appropriate more value from local R&D when it has stronger internal linkages across firm locations. Although there are many other concrete manifestations of internal linkages—from systematic meetings to enhanced communication mechanisms across clusters—our interviews with semiconductor industry managers identified a specific form of internal linkage that enhances appropriation: *cross-cluster teams*. Mapping internal linkages to concrete data on cross-cluster teams allows us to move our discussion of appropriation mechanisms beyond the abstract analysis most common in that literature.

2.4. Coordination Cost and the Selective Use of Internal Linkages

Cross-cluster teams need to overcome technological, organizational, and geographical barriers (Frost et al. 2002) and are costly to manage (Doz et al. 2006). Furthermore, although cross-cluster teams can enhance monitoring and control at the firm level, a lack of autonomy might dampen local researchers' incentive to innovate. Given the cost considerations, firms will only implement strong internal linkages when the benefits outweigh the costs, namely, when the threat of knowledge expropriation is high.

Note that a technology cluster with a large number of R&D-intensive firms in the same technological field is not necessarily a high-risk location. Firms face competition in both technology space and product market space, so collocating firms competing in the same technology space may not face business rivalry in the product market space. The multidimensional relationships among local entities (Cohen 1995, p. 230) allow us to separate appropriation incentives from knowledge-sourcing incentives. Collocated firms may share similar technological backgrounds—even engaging in patent races—without competing in the same product market. Industry-specific market information and other complementary resources reduce the risks associated with knowledge exchanges, allowing symbiotic relationships to develop. If internal linkages are purely mechanisms of knowledge sourcing, we should observe stronger internal linkages in clusters with a large number of firms in the same technological field (to access greater knowledge pools). If, on the other hand, internal linkages are a hedge against knowledge expropriation, we should observe internal linkages more frequently in clusters where neighbors share the same product market.

Therefore, we would expect to observe more cross-cluster teams at locations with higher expropriation risks, e.g., in clusters with a large number of direct competitors in the same product market. Furthermore, if cross-cluster teams enhance knowledge appropriation, we would expect two effects at the same time. On the one hand, because cross-cluster teams facilitate the integration of local innovations with complementary firm knowledge, we would expect to see more knowledge flowing across clusters in the presence of cross-cluster teams. On the other hand, because cross-cluster teams reflect internal interdependence, increase the firm-specific nature of projects (i.e., stronger complementarities with firm-specific resources), and thereby raise the learning barriers outsiders must overcome, we would expect to see less knowledge outflow to local competitors in the presence of cross-cluster teams.

It is important to note that we would also expect to see cross-cluster teams employed more intensively in clusters with more learning opportunities. In such cases, cross-cluster teams can facilitate the transfer of locally absorbed knowledge back to a firm's corporate center. Nevertheless, we believe that firms use cross-cluster teams as a strategic response to local expropriation risk *in addition to* using them as a response to knowledge-sourcing opportunities. Thus, we expect to find that cross-cluster teams are more prevalent when firms perceive higher expropriation risk in a cluster, given the learning opportunities in the region. The empirical analysis studies the dual roles played by cross-cluster teams in multilocation firms and attempts to distinguish a knowledge-sourcing effect from a knowledge-appropriation effect.

3. Empirical Design

Our empirical design focuses on the interactions between firms' technology strategies and product-market strategies, in three steps. First, we determine whether firms are more likely to use cross-cluster teams in clusters with large numbers of competitors. Second, we test whether innovations created by cross-cluster teams are used by other locations within the firm—an indication of internalization. Third, we determine whether innovations associated with cross-cluster teams are less used by collocated competitors—an evaluation of cross-cluster teams as a knowledge appropriation mechanism.

These three empirical analyses offer a thorough test of the propositions introduced in §2.3. In all cases, we calculate our variables using the same sample and data sources (§3.1) and the same definition of a cluster (§3.2). Most variables are used in all three steps (for example, the measurements that characterize the competitive environment of a cluster, as described in §3.5.1). However, there are differences among the steps in terms of dependent variables (described in §3.4) and the associated model-specific control variables (described in §3.5.3).

3.1. Sample

We chose the worldwide semiconductor industry in 1998–2001 as our empirical setting for several reasons. First, innovation is a key factor for success in semiconductors. Firms invest relentlessly in R&D to introduce new products and improve production processes (Stuart 2000). Moreover, semiconductor firms routinely patent their innovations, and patent data have been used to trace the traits and geographic distribution of innovation (Stuart 2000). Second, the benefit of knowledge transfer between firms has been shown to drive agglomeration in the industry (Saxenian 1994). High levels of geographic concentration also suggest that semiconductor firms have developed strategies to protect proprietary knowledge. Finally, semiconductors are a truly global industry. Firms vary widely by size and by number of locations, and often collocate with universities, national labs, and other industries (e.g., aerospace) conducting R&D in semiconductors. This heterogeneity allows us to identify how variations in local competitive environments affect firms' knowledge appropriation strategies and their organization of geographically dispersed R&D projects.

We assembled our data set from several sources. First, we identified innovating semiconductor firms using patent data from the Derwent World Patent Index (DWPI).[3] Because patent data include innovations that occur outside of R&D facilities, they are a more inclusive measure of innovation activity than, for example, the number of labs a firm runs or the amount it spends on R&D. Information from semiconductor patents applied for between 1998 and 2001, and granted between 2001 and 2003, results in a sample of 61,956 patents. Of these, 28,334 were granted in the United States.

Many of these patents are linked to the same innovation, either because patents were filed in multiple countries or because an application in a given country spun out multiple patents. Failing to recognize multiple patents associated with a single innovation could overestimate a location's innovation output (because a single innovation would be counted multiple times) or underestimate its backward and forward citations (because citations from or to a single innovation would be attributed to several patents). Thus, following Gittelman and Kogut (2003), we used Derwent families of patents as our unit of analysis. Each Derwent family encompasses patents granted in all countries that are identical in technology, inventors, and locations, but that differ in the scope of their claims. We restricted our sample to families

[3] We relied on Derwent's technological classification to obtain the universe of semiconductor patents. DWPI applies a consistent classification system to all patents. Classes used in this study are U11 (semiconductor materials and processes), U12 (discrete devices), U13 (integrated circuits), and U14 (memories, film, and hybrid circuits). For more details, see http://scientific.thomson.com/support/patents/dwpiref/reftools/classification.

with at least one American member and built forward and backward citation variables using only citations from and to American patents. (This avoids biases created by variations in citation standards and practices across legal jurisdictions.) The final sample consists of 23,383 patent families whose assignees are American and foreign firms, universities, governments, and industry-sponsored research labs. Patent families have an average of 2.6 foreign patent members and 1.2 American patent members. For the 624 patent families with more than one assignee, we considered all assignees (and not only the first one).

We supplemented our initial sample with directories of semiconductor plants, fabless companies, and the institutions responsible for scientific publications. Information about plants comes from the Strategic Marketing Association's quarterly data sets of World Fab Watch for 1998 to 2001. The data sets encompass manufacturing facilities for a wide range of products, including memories, microprocessors, and generic and specific chips.[4] We obtained information on fabless companies for the same period from the Gartner Group's annual Directory of Fabless Semiconductor Companies. To assess a location's scientific activities, we extracted from ISI Web of Knowledge all journal publications in the sample period that listed "semiconductor" or "semiconductors" as keywords. These sources provide a comprehensive map of the industry at multiple levels: innovation (23,383 patent families), production (974 plants), research (50,387 scientific publications), and development (549 fabless companies).

Because we treat multiunit firms as single integral entities, and because internal organization is a central concept of this study, we made additional efforts to identify the ultimate parent for every entity in our sample. First, for each year, we matched the patent assignees, plants, and fabless companies to firms in the corresponding Directory of Corporate Affiliations (DCA). Second, for organizations not identified in DCA, we searched Dun and Bradstreet's Million Dollar Database. Finally, we checked affiliation changes using SDC Platinum, company websites, and various industry publications. These steps allowed us to map 4,125 assignees in our sample to 2,217 unique organizations. We identified another 721 organizations among fabless and manufacturing firms that did not own patents.

Although we characterize local competitive environments using data for all organizations, we focused our analysis of R&D strategies on 16 innovating firms, or the top 1% number of assignees in the industry, as ranked by the number of patents they hold.[5] We did this because most of the semiconductor industry has features typical of an oligopoly industry; as a group, this top 1% of firms holds 50% of the total patents granted and 40% of the plants. With the cost of developing new chips and building new manufacturing plants running to the billions, there is a clear divide between industry leaders and everyone else. Therefore, we expect semiconductor industry leaders to have qualitatively different innovation strategies from the thousands of industry followers. Previous studies of the semiconductor industry used samples with a similar composition (Stuart and Podolny 1996, Ziedonis 2004). As part of our robustness checks, we replicated our analyses using an alternative sample composed of the top 30 firms—representing the top 5%—and obtained similar results.

3.2. Cluster Definition

Defining clusters is a crucial step in our empirical setup. Three elements must be specified: (1) the variable used to identify concentration levels of economic activity, (2) the geographic unit over which that variable is measured, and (3) the threshold level above which a concentration of firms is considered a cluster. Our definition of clusters departs from most previous work in these three dimensions.

First, we are interested in concentration of technological activity (technology clusters) instead of production density. Because agglomeration patterns for R&D and production differ (Audretsch and Feldman 1996), we followed Alcácer (2006) and used the geographic distributions of inventor activities instead of more conventional variables such as industry employment, plant output, or product sales.

Second, instead of relying upon predetermined administrative boundaries such as states, we applied a mathematical algorithm that uses latitude and longitude data to identify technology clusters. We did this for two reasons. First, there is no single administrative unit defined across all countries. We had to either focus on a specific country (e.g., the United States), which fails to capture important features of global firms, or use a mix of different geographic units (e.g., states in the United States, prefectures in Japan, and provinces in Europe), which may create unexpected country biases. Second, technology clusters do not necessarily follow predetermined administrative boundaries, which is clear after a quick inspection of inventor locations in, for example, the northeastern United States or central Japan. One administrative

[4] World Fab Watch data reports five types of manufacturing facilities: fabs, test facilities, assembly facilities, pilot fabs, and test fabs. We used only fabs—manufacturing facilities that were fully operative.

[5] The 16 firms are AMD, Intel, IBM, Texas Instruments, Hitachi, Matsushita, NEC, Siemens (including Infineon), Toshiba, Mitsubishi, Samsung, Micron, Fujitsu, TSMC, Hyundai, and STMicroelectronics.

unit may encompass multiple clusters, and one technological cluster may expand across several administrative lines.

Third, we defined clusters by the actual distribution of inventor locations following a three-step approach. (More detailed explanations and a comparison with alternative methods to define clusters are available upon request.) In the first step, we identified the location of each element in the sample (i.e., a patent inventor, plant, fabless company, or scientific publication) and matched them to two comprehensive sources of geographic names. For U.S. locations, we obtained latitude and longitude information for all 38,261 locations from the Geographic Names Information System of the U.S. Geological survey. For foreign locations, we used the Geonet Names Server (GNS) of the National Geospatial Intelligence Agency. Besides its wide coverage of 5.5 million location names worldwide, the GNS data set uses phonetic variations to capture spellings from different alphabets (as in Asian countries) and from alphabets with extra characters (as in Scandinavian and Slavic countries). Ambiguous matches were checked manually by native residents from various countries and areas. As a result, we were able to assign latitudes and longitudes to 61,385 of the 61,461 unique foreign locations in the original sample.

In the second step, we developed a mathematical algorithm to identify geographic clusters using latitude and longitude data. This step defined clusters not only by the geographic distance among locations—as in many traditional clustering methods—but also by the variations in inventor density in neighboring areas. For example, a rapid decrease in density might signal the end of a cluster, and a continuous level of density might signal a long or irregularly shaped cluster. Accordingly, the algorithm assigned two locations to the same cluster when there was a continuity of high-density locations between them, despite their geographic distance. In contrast, two locations that are not far apart geographically could be separated by a stretch of low-density areas and be identified as two distinct clusters. Our clustering algorithm offers the additional advantage of having the number of clusters emerge organically from the data instead of being set arbitrarily ex ante. This method produced 304 geographic units. Note that a patent family could be allocated to more than one cluster if its inventors were located in different clusters.

In the final step, we assigned plants, fabless companies, and publications to geographic units defined with patent data. In most cases, they fell within a cluster identified by the algorithm. In cases where locations fell outside of identified clusters, we calculated the shortest distance to an existing cluster. Locations that were less than 20 miles from an identified cluster are considered part of that cluster. Locations that were more than 20 miles from an identified cluster were clustered again with the same algorithm. Within the main sample, six and 31 geographic units were added by fabless and plant data, respectively.

Although the previous process generated 341 geographic units, not all of these units qualify as clusters. Numerous approaches have been used to determine whether a geographic concentration of economic activity is large enough to be categorized as a cluster. For example, Ellison and Glaeser (1997) identified clusters by visually inspecting a plot of their agglomeration measurement; Alcácer (2006) used the upper quartile of geographic patent concentration; and Delgado et al. (2010) defined clusters as those in the top 20% of their specialization-size measurement. For this study, we restricted our analyses to the top 25 locations, which together contained 84% of all patent families (88% of American patents), 57% of all plants, 67% of all fabless, and 76% of all publications.[6]

Table 1 shows descriptive statistics for the top 25 technological clusters used in our analysis, ranked by the average share of global patents from 1998 to 2001 (column (6)).[7] The first three columns contain the city, state, and country of the most common inventor location for a given cluster. Silicon Valley and the cluster around Tokyo were almost tied in terms of global shares of innovations in the sample period, followed by Poughkeepsie–Armonk (NY), Kobe in Japan, Boise (ID), and San Diego (CA). The United States and Japan had the largest number of clusters at 13 and 5, respectively. Italy, Germany, Singapore, South Korea, and Taiwan also hosted locations among the top 25 technological clusters.

Column (4) of Table 1 lists the best known city in a cluster, and column (5) shows the percentage of patents in a cluster that were assigned to the most common city—a figure that indicates how dense or dispersed a cluster is. For example, the most geographically concentrated technological cluster among the top 25 clusters was in Singapore—with 97% of inventors in Singapore City—followed by Boise (ID). Poughkeepsie–Armonk was the most dispersed, reflecting population dispersion in the metropolitan area of New York City. The technological clusters around San Diego (CA), Cambridge (MA), and Silver Spring (MD) were also among the most geographically dispersed.

Columns (7)–(9) introduce each cluster's global share of plants, fabless, and publications. Note that

[6] We replicated our analysis using all locations in which our sample firms had some presence. The results, available upon request, were similar to those obtained using the top 25 clusters only.

[7] These are the clusters in which the 16 firms in our sample operated.

Table 1 Characterization of Top 25 Clusters

(1) Main geographic unit	(2) State	(3) Country	(4) Best known city	(5) Patents in main city (%)	(6) Share of global patents (%)	(7) Share of global plants (%)	(8) Share of global fabless (%)	(9) Share of global publications (%)	(10) For-profit innovators (%)	(11) Large for-profit innovators (%)	(12) Universities (%)
San Jose	CA	US	San Francisco	20	15	6	38	1	95	69	65
Tokyo		JP		33	15	6	0	3	94	94	48
Wappingers Falls	NY	US	Poughkeepsie–Armonk	6	9	2	2	1	87	77	85
Hyogo Prefecture		JP	Kobe	16	5	4	0	1	96	95	43
Boise	ID	US		70	4	1	0	0	100	91	—
San Diego	CA	US		9	4	2	6	1	89	68	63
Seoul		KR		37	3	2	3	1	86	67	30
Hsinchu City		TW		40	3	6	12	1	94	62	13
Essex Junction	VT	US	Burlington	24	2	0	0	0	93	88	100
Plano	TX	US	Dallas–Fort Worth	26	2	2	1	0	91	82	89
Cambridge	MA	US	Boston	8	2	3	3	1	90	78	85
Portland	OR	US		22	2	2	1	0	93	87	100
Austin	TX	US		68	2	2	2	0	94	82	100
Nagoya		JP		12	2	3	0	1	96	95	75
Munich		DE		47	2	1	0	0	84	100	13
Chandler	AZ	US	Phoenix	26	1	3	1	0	93	85	100
Taipei		TW		49	1	0	5	0	89	65	33
Bergamo		IT	Milan	11	1	1	0	0	76	92	50
Colorado Springs	CO	US	Denver	31	1	1	2	0	90	83	36
Minneapolis	MN	US		10	1	1	0	0	90	78	78
Silver Spring	MD	US	Washington DC	9	1	1	1	1	75	65	43
Suwa region		JP	Nagano	27	1	2	0	0	100	97	—
Busan		KR		23	1	1	0	0	72	73	10
Hitachi		JP		38	1	1	0	1	87	98	0
Singapore		SG		97	0	2	0	0	76	95	92

the global plant share for the two top technological clusters was three times lower than their global innovation share, suggesting that, as reported previously by Alcácer (2006) and Audretsch and Feldman (1996), agglomeration based on production and R&D did not overlap. Hsinchu City in Taiwan and Phoenix (AZ) also support this claim; in both cases the global plant share was double that of global innovation centers, suggesting that both clusters emphasized production. Fabless were concentrated in Silicon Valley and, to a much smaller degree, in Hsinchu City. Publications were very dispersed geographically, with the Tokyo area accounting for only 3% of the total.

Columns (10)–(12) show the composition of clusters in terms of their share of for-profit innovators, large for-profit innovators, and universities. For-profit organizations constituted the majority of innovators in technological clusters with the exceptions of clusters in Washington (DC)—where innovations by government agencies are numerous—and Singapore—where innovations by nonprofit groups are common. Low shares of large for-profit innovators demonstrate the entrepreneurial spirit of Silicon Valley, San Diego, Taipei, and Hsinchu City. At the other extreme, Munich, Hitachi, and Nagano show high concentration of large for-profit firms. Universities were the main group of nonprofit innovators in Burlington (VT), Portland (OR), and Phoenix (AZ) and played no role in Boise or Busan.

Taking the figures together, Table 1 shows that innovative clusters may not include large concentrations of production, and that the composition of local competitive environments will vary by cluster.

3.3. Models

To identify firms' strategic organization of R&D projects across clusters, we compare the technologies developed in different local competitive environments, controlling for firm and patent characteristics. Specifically, we explore three dimensions of innovations: whether the innovation was associated with inventors located across clusters, whether this type of innovation was internalized (cited as prior art by innovations developed in other clusters), and whether the innovation was cited less often by competitors in the same cluster. Although the models estimated in each dimension differ in terms of dependent variables and some independent variables, they all follow the same general structure of

$$DepVar^{(k)}_{fict} = f\big(\beta^{(k)}_1 C_{ict} + \beta^{(k)}_2 X_f + \beta^{(k)}_3 Y_{ict} + \beta^{(k)}_4 Z_{ct} + \zeta_t + \tau_{ctry} + \gamma_{tech} + \upsilon_i + \varepsilon_{fict}\big), \quad (1)$$

where $k = 1, 2, 3$ indicates the three sets of analyses described in §3.4. Accordingly, $DepVar_{fict}$ is one of the three variables—$cross\_cluster_{fict}$, $cross\_cluster\_self\_citation_{fict}$, and $local\_citation\_by\_competitor_{fict}$—for steps 1, 2, and 3, respectively; C_{ict} is a vector of cluster-specific variables capturing the competitive environment faced by firm i in cluster c and year t (§3.5.1); X_f is a vector of patent family-specific variables (§3.5.2); Y_{ict} is a vector of firm-specific variables characterizing firm i in cluster c and year t (§3.5.3); Z_{ct} is a vector of location characteristics in year t (described in §3.5.1); ζ_t, τ_{ctry}, γ_{tech}, and υ_i, are four sets of dummy variables for year, country, technology and firm fixed effects, respectively; and ε_{ict} is the error term.

The estimation technique varies according the nature of the dependent variable. We used logit estimation techniques for models with $cross\_cluster_{fict}$ as a dependent variable, given its binary nature. Because $cross\_cluster\_self\_citation_{fict}$ and $local\_citation\_by\_competitor_{fict}$ are count variables, we used negative binomial estimation techniques.[8]

3.4. Dependent Variables

3.4.1. Cross-Cluster Teams. Geographically dispersed R&D in a multilocation firm makes it more difficult for local competitors to access the technological know-how residing in a firm's other locations, which in turn minimizes knowledge expropriation. Teams spanning multiple clusters can also facilitate the transfer of local know-how throughout the organization (Lahiri 2010). Thus, we defined $cross\_cluster_{fict}$ as a binary variable equal to 1 if the patent family has at least one inventor in any of the focal clusters (top 25 clusters) and at least one inventor somewhere else (either in another focal cluster or in a location not included in the top 25 clusters).

3.4.2. Cross-Cluster Self-Citations. A key concept in this study is the extent to which an innovation creates value for the innovating firm. Although there is no direct measure of value, technologies highly dependent on internal resources are more likely to be utilized and further developed within the firm. Trajtenberg et al. (1997, p. 29) proposed using self-citations, defined as "the percentage of citing patents issued to the same assignee as that of the originating patent," to measure the "fraction of the benefits captured by the original inventor." Hall et al. (2005) have also suggested that citations to patents belonging to the same firm represent internalized knowledge transfers, bolstering the firm's competitive advantage. Hence, we used forward self-citations as a proxy for the value new technologies bring a firm. Specifically, we defined the variable $cross\_cluster\_self\_citation_{fict}$ as the number of self-citations received by patent family p by other patents whose assignee is firm i but which were generated in a cluster other than cluster c.[9] Because we are interested in firms as integrated organizations, citations among affiliated organizations were considered self-citations.

3.4.3. Local Citations by Competitors. If using cross-cluster teams leads to internalization and better knowledge appropriation, we should observe that patents associated with cross-cluster teams were less likely to be cited by competitors in the originating cluster. To test this proposition, we created a set of variables that capture the number of citations made by competitors located in the same cluster. Because our characterization of a cluster's competitive environment is diverse, the count of local citations by competitor varies accordingly. We defined $local\_citations\_by\_innovators_{fct}$ as the total number of citations to members in a family made by other assignees—regardless of their type; $local\_citations\_by\_profit_{fct}$ as the count of citations made by assignees classified as for-profit organizations; $local\_citations\_by\_industry_{fct}$ as the count of citations made by assignees in firm i's industry; $local\_citations\_by\_segment_{fct}$ as the count of citations made by assignees in firm i's product segment; and $local\_citations\_by\_competitor_{fct}$ as the count of citations made by assignees identified as market competitors of firm i. Only citations made by patents originated in cluster c were counted. Section 3.5.1 describes how we classified assignees into profit, industry, segment, and market competitor.

Two issues related to citation-based measurements are worth exploring further. First, citation measures capture both the intensity and the speed of citations. Because our observation window ends in September 2010, any citations that occurred after that date were not included in the sample. Jaffe and Trajtenberg (2005) have shown that the lag of forward citations peaks at around five years. The period to accumulate citations in our sample ranges from seven to nine years, so our citations should represent the bulk of citations ultimately received.[10] We also included time fixed effects (ζ_t) to account for variations in citations across cohorts and to control for year-specific events.

Second, a common critique of citation-based measurements is the role of patent examiners. Recent research reveals that examiner citations account for 66% of all citations in an average patent, which may bias empirical tests (Alcácer and Gittelman 2006).

[8] We evaluated the use of negative binomial versus Poisson models through a set of Hausman tests. Results of these tests favor negative binomial models and are available upon request.

[9] We aggregated cross-cluster self-citations received by any member of family patent p.

[10] We plotted the stream of citations for patents in our sample and found that, in fact, citations peaked before September 2010.

To avoid this problem, our main models were estimated using citations listed by assignees only. In our sample, about 35% of the patent families that received at least one inventor citation also had at least one self-citation. The number was 42% when both inventor and examiner citations were considered. For robustness checks, we repeated our analysis using all citations to a patent regardless of their source.

3.5. Independent Variables

3.5.1. Characterizing Clusters (C_{ict}, Z_{ct})**.**
Competition (C_{ict}). Firms sharing the same technological space have better absorptive capacity for each other's knowledge (Cohen and Levinthal 1990), so the likelihood of knowledge expropriation is larger in areas with many firms doing the same type of R&D. Meanwhile, knowledge outflows only create high risks when the recipients of that knowledge target the same product market (Dushnitsky and Shaver 2009). Therefore, we follow two dimensions—technology space and product market—to characterize the competitive environment at the cluster-year level.

Along the technology space, competitors are organizations that innovate in the semiconductor field. The variable *innovators* represents the number of unique assignees with semiconductor patents in a given cluster year. We classified assignees into two groups, *innovators_profit* and *innovators_nonprofit*, to capture the number of for-profit and nonprofit assignees, respectively. In addition, we used the status information on patent applications to further classify *for-profit* assignees into small or large entities, thus creating the variables *small_innovators* and *large_innovators*. In the case of nonprofit assignees, we classified them manually into three groups: universities (*universities*), government agencies (*government*), and other nonprofits (*other_nonprofit*).

Along the second dimension, we defined competitors as firms that share the same product market. For every focal firm in our sample, we used Hoover's Online to identify the industry (four-digit SICs), market segments within semiconductors (e.g., memory chips and modules, microprocessors, etc), and the names of direct competitors. Then we counted the number of for-profit assignees in the same industry (*in_industry* and *not_in_industry*), in the same market segment (*in_segment* and *not_in_segment*), or on the list of direct competitors (*competitors* and *not_competitors*). The self-reported competition data from Hoover's serve our purpose well because managers make strategic decisions based upon perceived competition in a technology cluster. These variables vary by firm, cluster, and year.

Other cluster variables (Z_{ct}). We completed our characterization of local innovation environments with three more variables—*plants_in_cluster*, *fabless_in_cluster*, and *publications_in_cluster*—which represent the numbers of plants, fabless companies, and publications per cluster year, respectively.

Zhao (2006) has shown that internalization varies according to the property rights regime of a given country. Thus, we also controlled for variations in country-specific intellectual property right regimes by a set of country dummies τ_{ctry}.

3.5.2. Characterizing Patent Families (X_f)**.** As discussed in §2.3, cross-cluster teams and internalization can be associated with knowledge sourcing. To control for this, we created a set of dummy variables that indicate whether any member of the focal patent family cites prior art granted to assignees in the same cluster.[11] These variables are defined by the competition measurement used in a specific model: $cites\_local_{fc}$ is equal to 1 if any member of the family cites at least one previous local patent (regardless of the assignee type of such patent);[12] $cites\_local\_profit_{fc}$ is equal to 1 for families with at least one patent member citing an assignee classified as a for-profit entity; $cites\_local\_industry_{fc}$ is equal to 1 if the citation was to assignees in the semiconductor industry; $cites\_local\_segment_{fc}$ is equal to 1 if there were citations to assignees in the same segment; and $cites\_local\_competitor_{fc}$ is equal to 1 if there were citations to market competitors.

Patents associated with cross-cluster teams that are more internalized or are less cited by proximate competitors may also have low intrinsic values. Through all specifications we controlled for innovation quality through $claims_f$—the summation of patent claims across all American members of a patent family (Lanjouw and Schankerman 2004).[13] Although our sample was drawn by sampling by technology, our results may also be driven by technological differences within semiconductors. We therefore added a set of dummy variables by technology classes (γ_{tech}).

Finally, some patent-family variables are relevant only for specific analyses. For example, in step 2—testing for cross-cluster self-citations—we controlled for the baseline propensity that a patent is more self-cited because of the sheer number of subsequent patents granted to a firm. Specifically, we constructed a patent-variant variable $patent\_stock_{fit}$ to represent the number of patents that firm i obtained between the time the first patent within a focal firm was granted until September 2010. In steps 2 and 3, we used

[11] We are very thankful to an anonymous reviewer who suggested this approach to control for knowledge sourcing.

[12] This variable indicates citations to any patent.

[13] We used an alternative measure for innovation quality—total number of countries where an innovation has been patented (Putnam 1996)—and obtained very similar results.

$total\_citations_{ft}$—defined as the total number of citations received by American patents in family f—as the exposure variable for negative binomial models. Because the coefficients for exposure variables are forced to be 1, we are essentially estimating the ratios of cross-cluster self-citations to total citations in step 2, and the ratio of citations by local competitors to total citations in step 3.

3.5.3. Characterizing Firms (Y_{ict}). Technologies closely linked to manufacturing or product design may have different characteristics from others. In addition to the variables measuring local competitive environments (as described in §3.5.1), we used two dummy variables, $with\_plant_{ict}$ and $with\_fabless_{ict}$, to indicate whether a particular firm has plants or fabless units in cluster c and year t. Both variables vary by firm-cluster-year and are used in steps 1 and 2. Finally, Alcácer et al. (2009) showed that firms' prior-art citation practices vary widely. We added firm fixed effects (v_i) to control for these differences and any time-invariant firm characteristics.

Table 2 Summary Statistics

Variable	*N*	Mean	SD	Min	Max
Cluster variables					
innovators	7,730	65.810	66.518	4	253
innovators_profit	7,730	61.156	63.148	3	240
large_innovators	7,730	49.206	45.567	2	153
small_innovators	7,730	11.950	21.210	0	90
innovators_non_profit	7,730	4.653	4.398	0	16
universities	7,730	2.928	3.675	0	14
government	7,730	1.531	1.720	0	6
other_nonprofit	7,730	0.194	0.484	0	2
plants_in_cluster	7,730	22.291	18.685	1	54
fabless_in_cluster	7,730	25.045	61.122	0	206
publications_in_cluster	7,730	268.290	255.117	1	754
Firm-cluster variables					
in_industry	7,730	19.014	19.692	0	73
not_in_industry	7,730	42.142	43.897	2	168
in_segment	7,730	12.794	14.046	0	65
not_in_segment	7,730	48.362	50.091	2	213
competitors	7,730	6.928	6.200	0	27
no_competitors	7,730	54.228	59.451	2	234
with_plant	7,730	0.162	0.368	0	1
with_fabless	7,730	0.021	0.142	0	1
Patent variables					
cross_cluster_team	7,730	0.261	0.439	0	1
cross_cluster_self_citations	7,730	1.225	5.680	0	137
local_citations_by_innovators	7,730	2.355	7.928	0	188
local_citations_by_profit	7,730	2.355	7.928	0	188
local_citations_by_industry	7,730	0.439	3.665	0	188
local_citations_by_segment	7,730	0.432	3.730	0	188
local_citations_by_competitor	7,730	0.187	0.985	0	37
patent_stock	7,730	17,378	9,046	4,304	42,096
total_citations	7,730	8.803	16.089	1	201
cites_local	7,730	0.352	0.478	0	1
cites_local_profit	7,730	0.352	0.478	0	1
cites_local_industry	7,730	0.101	0.301	0	1
cites_local_segment	7,730	0.090	0.286	0	1
cites_local_competitor	7,730	0.069	0.254	0	1
claims	7,730	21.032	18.654	1	521

Table 2 shows the summary statistics for our variables. The final number of observations, 7,730, represents patent families associated with the top 16 innovators in the industry. These families had at least one cross-cluster self-citation and were developed in one of the top 25 technological clusters.

4. Empirical Results

4.1. Are Firms More Likely to Use Cross-Cluster Teams When There Are More Competitors Nearby?

The first step in our analysis was to determine if innovations generated by cross-cluster-teams were more likely to emerge in clusters with higher levels of competitor presence. The model to estimate was

$$cross\_cluster_{fict} = \beta_1^{(1)} C_{ict} + \beta_2^{(1)} X_f + \beta_3^{(1)} Y_{ict} + \beta_4^{(1)} Z_{ct} + \zeta_t + \tau_{ctry} + \gamma_{tech} + v_i + \varepsilon_{fict}. \quad (2)$$

Table 3 shows the results of estimating Equation (2) using logit. Standard errors were clustered by geographic cluster.

The positive coefficients of *innovators_profit*, *large_firms*, *in_industry*, *in_segment* and *competitors* suggest that the presence of competing organizations increases the tendency to use cross-cluster teams. For example, according to column (7), an extra product market competitor in a cluster increases the likelihood of using a cross-cluster team by 42%.[14] Note that some innovations are codeveloped in high-risk clusters and low-risk clusters. Such observations work against finding these results because both low- and high-risk clusters are associated with the presence of a cross-cluster team.[15] Finally, the number of nonprofit innovators has no effect on the use of cross-cluster teams.

Evidence that firms in our sample use cross-cluster teams to channel local knowledge acquired in the cluster to distant firm locations is mixed. The dummy variable for backward citations is positive and significant when it implies citations to any type of assignee, but not when it implies citations to competitors or firms in the same segment or industry. In other words, knowledge sourcing from market competitors does not trigger the use of cross-cluster teams, indicating that the "reverse knowledge integration" suggested by Frost and Zhou (2005), if it occurs, is sourcing mostly from noncompetitive institutions such as universities, research institutes, and innovators that are not direct competitors.

[14] The effect is calculated by keeping all other independent variables at their mean value. The marginal effect is 0.11 and the underlying probability, at mean values, is 0.26.

[15] Thanks to an anonymous reviewer from bringing this point to our attention.

Table 3 Logit Estimates on Cross-Cluster Teams

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Cluster and firm-cluster competitive environment variables							
innovators		0.007					
		(1.72)†					
innovators_profit			0.073				
			(2.79)**				
large_innovators				0.074			
				(2.49)*			
small_innovators				0.003			
				(0.31)			
in_industry					0.078		
					(2.74)**		
not_in_industry					−0.010		
					(1.40)		
in_segment						0.070	
						(2.82)**	
not_in_segment						−0.025	
						(1.66)†	
competitors							0.070
							(2.79)**
no_competitors							−0.003
							(0.84)
innovators_non_profit			−0.001		0.030	−0.053	−0.028
			(0.31)		(0.39)	(0.66)	(0.25)
universities				0.062			
				(0.62)			
government				−0.003			
				(0.29)			
other_nonprofit				0.155			
				(0.97)			
Cluster variables							
plants_in_cluster	0.158	0.102	0.165	0.174	0.114	0.084	−0.199
	(0.44)	(0.28)	(0.42)	(0.43)	(0.29)	(0.18)	(0.42)
fabless_in_cluster	0.003	−0.002	0.001	0.001	−0.003	0.003	−0.0001
	(0.88)	(0.44)	(0.33)	(0.14)	(0.89)	(0.68)	(0.04)
publications_in_cluster	0.001	0.001	−0.0002	−0.0001	−0.0003	−0.0004	0.0000000
	(1.10)	(0.56)	(0.15)	(0.08)	(0.33)	(0.34)	(0.02)
Firm-cluster variables							
with_plant	2.241	2.188	2.145	2.145	2.069	2.148	2.196
	(5.90)**	(6.04)**	(6.02)**	(5.89)**	(6.08)**	(6.32)**	(6.46)**
with_fabless	1.472	1.418	1.477	1.499	1.466	1.646	1.767
	(0.64)	(0.67)	(0.69)	(0.68)	(0.70)	(0.68)	(0.70)
Patent variables							
cites_local[a]	0.315	0.267	0.069	0.070	0.208	0.110	0.162
	(2.93)**	(2.94)**	(0.48)	(0.49)	(1.56)	(1.10)	(1.29)
claims	0.009	0.009	0.009	0.009	0.008	0.008	0.009
	(2.09)*	(2.13)*	(2.11)*	(2.11)*	(2.03)*	(2.13)*	(2.10)*
Constant	−1.188	−1.239	−1.244	−1.247	−1.383	−1.014	−0.991
	(3.39)**	(3.40)**	(3.46)**	(3.57)**	(3.94)**	(2.21)*	(2.77)**
Country effects	Y	Y	Y	Y	Y	Y	Y
Firm effects	Y	Y	Y	Y	Y	Y	Y
Technology effects	Y	Y	Y	Y	Y	Y	Y
Year effects	Y	Y	Y	Y	Y	Y	Y
Observations	7,730	7,730	7,730	7,730	7,730	7,730	7,730
Log pseudolikelihood	−3,712	−3,706	−3,697	−3,696	−3,689	−3,688	−3,689

Notes. Dependent variable is *cross_cluster_team*, defined as a binary variable equal to 1 for patent families with cross-cluster teams, 0 otherwise. Robust *z*-statistics are in parentheses. Standard errors have been clustered by geographic cluster.

[a] *cites_local* is defined differently across models to match the characterization of competition in the cluster: it is *cites_local* for models (1) and (2); *cites_local_profit* for models (3) and (4); *cites_local_industry* for model (5); *cites_local_segment* for model (6); and *cites_local_competitor* for model (7).

†Significant at 10%; *significant at 5%; **significant at 1%.

Two other variables deserve further comment. Having a plant in a cluster increases dramatically the chances of having a cross-cluster team. This may suggest that innovations in manufacturing require more coordination across production sites or specialized knowledge that resides in central R&D labs. Higher quality innovations—innovations with more claims—are also more likely to be associated with cross-cluster teams. Given the higher level of coordination that cross-cluster teams require, it is not surprising that firms are more inclined to use them when the potential benefits are higher.

Together, these findings suggest that high levels of competitors in a cluster, rather than proximity to other types of innovators, are linked to cross-cluster teams, and that cross-cluster teams appear to be associated with appropriation rather than with knowledge sourcing.

4.2. Are Innovations Created By Cross-Cluster Teams More Likely to Be Used at Other Locations Within the Firm?

The second step in our analysis investigated if innovations generated by cross-cluster teams are more likely to be cited by innovations created by the same firm in other clusters. The model to estimate was

$$\begin{aligned} &cross\_cluster\_self\_citation_{fict}\\ &\quad = \beta_1^{(2)} C_{ict} + \beta_2^{(2)} X_f + \beta_3^{(2)} Y_{ict} + \beta_4^{(2)} Z_{ct}\\ &\qquad + \zeta_t + \tau_{ctry} + \gamma_{tech} + \upsilon_i + \varepsilon_{fict}. \end{aligned} \tag{3}$$

Table 4 shows the results of estimating Equation (3) using negative binomial with $total\_citations_{ft}$ as the exposure variable. Because the dependent variable is the number of self-citations received by the focal patent, and the exposure variable is the total number of forward citations, we are essentially examining the patent's cross-cluster self-citation ratio. Ordinary least squares regressions with self-citation ratio as the dependent variable produced very consistent results. Standard errors were clustered by geographic cluster. (Note that $cross\_cluster_{fict}$ becomes an independent variable.)

Across specifications, innovations associated with cross-cluster teams are more internalized. For example, model (7) in Table 4 indicates that cross-cluster team patents receive 0.5 more cross-cluster self-citations, which is 40% of the average cross-cluster self-citations in the sample. The result is robust in magnitude and significance regardless of how competitors in a cluster are defined.

The total number of innovators in the cluster does not seem to have a significant impact on internalization. The effect of local competition emerges when firms compete in the product market but not when entities employ similar technology in different markets. Across various specifications of the local competitive environment, the coefficient on the number of local competitors is positive and significant. The more market competitors in a cluster, the more likely firms will self-cite patents they developed in that cluster. Note that this effect is *after* controlling for innovations developed by cross-cluster teams; therefore, other mechanisms that enhance internalization—mechanisms that we do not measure directly—must be at play. If self-citations proxy for internalized value, this finding suggests that, in highly competitive environments, firms are more likely to share technology development across the firm. Meanwhile, the presence of nonprofit innovators has little impact on the degree of internalization.

Together, these findings suggest that firms change the type of innovation they perform in a given location—beyond using cross-cluster teams—depending on the local competitive environment. Innovation produced in clusters with high direct competition is more intertwined with the firm's internal knowledge.

Note that there is little empirical support for the argument that internalization across clusters is driven by knowledge sourcing; i.e., families that draw from assignees' local innovation are *not* more likely to generate cross-cluster self-citations. Not surprisingly, the coefficient for *patent_stock* is positive and significant—the larger the pool of patents, the more likely that later citations are made to the focal patent. The coefficient of *with_plant* is also positive and significant, indicating that technologies closely linked to manufacturing processes are more firm-specific. Finally, note that the high ratios of cross-cluster self-citations in competitive clusters are not caused by the low intrinsic values (small denominators) of these patents; we failed to reject the hypothesis that the coefficient for claims is not distinct from zero.

4.3. Are Innovations Created by Cross-Cluster Teams Less Used by Competitors?

Our final step was to evaluate cross-cluster teams as mechanisms for appropriating knowledge. We did this by determining whether those innovations associated with cross-cluster teams are less cited locally. Specifically, we estimated the following equation:

$$\begin{aligned} &local\_citations\_by\_ENTITY_{fct}\\ &\quad = \beta_1^{(3)} C_{ict} + \beta_2^{(3)} X_f + \zeta_t + \tau_{ctry} + \gamma_{tech} + \upsilon_i + \varepsilon_{fict}. \end{aligned} \tag{4}$$

Table 5 shows the results of estimating Equation (4) using negative binomial, with $total\_citations_{ft}$ as the exposure variable. Thus, we are essentially estimating the ratio of local citations to total citations to the focal patent. Standard errors were clustered by geographic

Table 4 Negative Binomial on Cross-Cluster Self-Citations

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Cluster and firm-cluster competitive environment variables							
innovators		0.0003					
		(0.15)					
innovators_profit			0.004				
			(1.86)†				
large_innovators				0.008			
				(1.41)			
small_innovators				−0.003			
				(0.5)			
in_industry					0.035		
					(2.95)**		
not_in_industry					−0.006		
					(1.28)		
in_segment						0.029	
						(5.33)**	
not_in_segment						0.001	
						(0.31)	
competitors							0.030
							(2.25)*
no_competitors							−0.003
							(1.22)
innovators_non_profit			0.042		0.045	0.038	0.040
			(1.53)		(1.64)	(1.46)	(1.56)
universities				0.033			
				(1.31)			
government				−0.022			
				(0.18)			
other_nonprofit				−0.064			
				(0.93)			
Cluster variables							
plants_in_cluster	−0.750	−0.730	−1.020	−1.330	−0.990	−0.880	−1.050
	(1.5772)	(1.5638)	(1.5352)	(1.1345)	(1.4606)	(1.6662)	(1.5885)
fabless_in_cluster	0.076	0.072	0.105	−0.008	0.099	0.119	0.009
	(0.15)	(0.14)	(0.2)	(0.01)	(0.21)	(0.3)	(0.02)
publications_in_cluster	0.001	0.001	−0.0001	0.0002	−0.0002	−0.0003	0.0000
	(0.79)	(0.65)	(0.06)	(0.25)	(0.21)	(0.32)	(0.01)
Firm-cluster variables							
with_plant	0.935	0.933	0.902	0.891	0.846	0.879	0.920
	(6.72)**	(6.92)**	(6.18)**	(6.45)**	(5.87)**	(6.59)**	(6.66)**
with_fabless	−0.275	−0.276	−0.250	−0.225	−0.240	−0.066	−0.160
	(0.38)	(0.38)	(0.34)	(0.3)	(0.32)	(0.08)	(0.19)
Patent variables							
cross_cluster_team	0.518	0.518	0.506	0.503	0.496	0.492	0.499
	(4.56)**	(4.54)**	(4.61)**	(4.63)**	(4.59)**	(4.58)**	(4.56)**
cites_local[a]	0.270	0.227	0.236	0.235	0.170	0.107	0.133
	(1.67)†	(1.66)†	(1.51)	(1.79)†	(1.45)	(0.93)	(1.16)
claims	0.002	0.002	0.002	0.002	0.002	0.002	0.002
	(0.94)	(0.94)	(0.96)	(0.95)	(0.89)	(1.04)	(0.97)
patent_stock	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
	(1.94)†	(1.78)†	(2.02)*	(1.72)†	(2.06)*	(2.32)*	(2.19)*
Constant	−2.690	−2.695	−2.703	−2.766	−2.733	−2.424	−2.577
	(5.23)**	(5.29)**	(5.22)**	(5.26)**	(5.67)**	(5.72)**	(5.32)**
Country effects	Y	Y	Y	Y	Y	Y	Y
Firm effects	Y	Y	Y	Y	Y	Y	Y
Technology effects	Y	Y	Y	Y	Y	Y	Y
Year effects	Y	Y	Y	Y	Y	Y	Y
Observations	7,730	7,730	7,730	7,730	7,730	7,730	7,730
Log pseudolikelihood	−6,263	−6,293	−6,289	−6,281	−6,285	−6,278	−6,285

Notes. Dependent variable is *cross_cluster_self_citations*. Robust z-statistics are in parentheses. Standard errors have been clustered by geographic cluster.

[a] *cites_local* is defined differently across models to match the characterization of competition in the cluster: it is *cites_local* for models (1) and (2); *cites_local_profit* for models (3) and (4); *cites_local_industry* for model (5); *cites_local_segment* for model (6); and *cites_local_competitor* for model (7).

†Significant at 10%; *significant at 5%; **significant at 1%.

Table 5 Negative Binomial on Local Citations by Competitors

	Dependent variable: *local_citations_by_*						
	innovators		*profit*		*industry*	*segment*	*competitor*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Cluster and firm-cluster competitive environment variables							
innovators		0.117					
		(3.56)**					
innovators_profit			0.120				
			(5.14)**				
large_innovators				0.134			
				(2.06)*			
small_innovators				0.016			
				(0.63)			
in_industry					0.017		
					(3.41)**		
not_in_industry					0.003		
					(0.2)		
in_segment						0.011	
						(2.22)*	
not_in_segment						0.008	
						(1.17)	
competitors							0.022
							(0.29)
no_competitors							−0.001
							(0.23)
innovators_non_profit			0.018		0.002	0.002	0.002
			(0.92)		(1.25)	(0.85)	(0.09)
universities				0.010			
				(0.57)			
government				0.035			
				(0.8)			
other_nonprofit				0.071			
				(1.46)			
Patent variables							
cross_cluster_team	−0.544	−0.549	−0.556	−0.552	−0.507	−0.569	−0.561
	(4.10)**	(4.17)**	(4.05)**	(4.12)**	(2.28)*	(2.99)**	(3.60)**
cites_local[a]	0.401	0.391	0.203	0.219	0.571	0.890	0.709
	(8.25)**	(8.18)**	(0.56)	(0.64)	(0.84)	0.3481	(1.41)
claims	0.010	0.010	0.009	0.009	0.014	0.016	0.012
	(2.92)**	(2.91)**	(2.93)**	(2.85)**	(4.16)**	(5.97)**	(3.63)**
Constant	−2.118	−2.362	−2.346	−2.456	−5.019	−4.947	−5.081
	(13.66)**	(9.78)**	(10.04)**	(10.24)**	(8.89)**	(7.90)**	(5.08)**
Country effects	Y	Y	Y	Y	Y	Y	Y
Firm effects	Y	Y	Y	Y	Y	Y	Y
Technology effects	Y	Y	Y	Y	Y	Y	Y
Year effects	Y	Y	Y	Y	Y	Y	Y
Observations	7,730	7,730	7,730	7,730	7,730	7,730	7,730
Log pseudolikelihood	−11,696	−11,678	−11,674	−11,647	−4,004	−4,085	−2,848

Notes. Dependent variable is defined as citations made locally by competitors. The dependent variable varies to match the characterization of competition in the cluster: it is *local_citations_by_innovators* for models (1) and (2); *local_citations_by_profit* for models (3) and (4); *local_citations_by_industry* for model (5); *local_citations_by_segment* for model (6); and *local_citations_by_competitor* for model (7). Robust z-statistics are in parentheses. Standard errors have been clustered by geographic cluster.

[a] *cites_local* is defined differently across models to match the characterization of competition in the cluster: it is *cites_local* for models (1) and (2); *cites_local_profit* for models (3) and (4); *cites_local_industry* for model (5); *cites_local_segment* for model (6); and *cites_local_competitor* for model (7).

*Significant at 5%; **significant at 1%.

cluster. Three issues in this table require further discussion. First, as in Table 4, $cross\_cluster_{fict}$ is an independent variable that characterized a patent family. Second, the dependent variable varies across columns depending on how the local environment is characterized. In columns (1) and (2), the dependent variable is $local\_citations\_by\_innovators_{fct}$; in columns (3) and (4), it is $local\_citations\_by\_profit_{fct}$; in column (5), it is $local\_citations\_by\_industry_{fct}$; in column (6), it is $local\_citations\_by\_segment_{fct}$; and in column (7), it is $local\_citations\_by\_competitor_{fct}$. Although all variables represent the same concept—citation by assignees that may pose a threat—the definition of assignees varies to match the characterization of the cluster's competitive environment.

Innovations associated with cross-cluster teams are cited less often by competitors' local innovations. For example, for model (7) in Table 5—our more precise measurement of competition—innovations associated with cross-cluster teams receive 0.56 less local citations by competitors than innovations that are not associated with cross-cluster teams. This finding supports our hypothesis that cross-cluster teams are an effective appropriation mechanism and are not solely associated with learning and knowledge sourcing. (Note that this result is after controlling for knowledge sourcing through the backward citation variables, whose coefficients are positive and significant in models (1) and (2).)

The more assignees of a given type (*innovators, innovators_profit, in_industry, in_segment, competitors*) in a cluster, the more patents they might obtain there and the more likely those patents would cite the focal patent families in our sample. As such, positive coefficients for these variables would suggest a size effect, whereas negative coefficients would indicate that appropriation mechanisms are so effective that citations by competitors decrease *even though* the total number of citations increases. As expected, most coefficients are positive and statistically significant—a sheer size effect. However, we note that in model (7)—a model with the most precise competition measurement—the coefficient is not significant, suggesting that appropriation mechanisms beyond cross-clusters teams may be at play. That these alternative mechanisms were not measured explicitly suggests more research is required.

In terms of control variables, more valuable innovations are more likely to be cited by other assignees in the cluster, whether those assignees are competitors or not.

5. Robustness checks

Our findings were consistent with our hypothesis that R&D projects in competitive clusters are more internalized across clusters, and are less often expropriated by nearby competitors. Next, we conducted a series of robustness tests using alternative samples and estimation techniques, and with our variables defined differently.

First, there may be significant correlations among the variables we used to characterize clusters, namely the number of plants, fabless, publications, and assignees. Multicollinearity increases standard errors and makes it less likely that coefficients will be statistically significant. In other words, strong correlations among the cluster variables would work *against* us finding any significance for the local competitive effect, which we obtained anyway. To check this, we reestimated the models in Tables 3–5 in four different ways to minimize any effects from collinearity. First, when we orthogonalized variables in C_{ict} and Z_{ct} for each model, results were stronger in terms of significance and magnitude.[16] Second, we used one variable to characterize the cluster in each model (*innovators, innovators_profit, in_industry, in_segment, competitors* for models (2)–(7), respectively). Our results were similar, though because each variable encompassed multiple cluster characteristics, their interpretation became less clear. Third, we used cluster characteristic ratios instead of individual variables. For example, instead of using *competitors, not_competitors* and *innovators_nonprofit* for model (7) in any table, we created $ratio\_competitor = competitors/not\_competitors$ and $ratio\_nonprofit = innovators\_nonprofit/innovators$. This change eliminated correlation problems. The results were similar, though again the coefficients are harder to interpret. Fourth, we repeated our analysis with all locations identified in §3.2 rather than using the top 25 clusters alone. Including more geographic units increased heterogeneity across clusters. As a result, the correlation between plants and competitors decreased, presumably because the sample included more geographic units rich in manufacturing facilities but poor in innovation. In all instances, however, the alternative results did not depart significantly from those presented in §4.

Second, we reestimated all models using a different method to define clusters: hierarchical clustering with centroid linkages. This method began with each location as a separate group. The two clusters with the shortest Euclidian distance were combined into one, whose new geographic coordinates were the mean longitude and latitude of all locations in the group. This process was repeated until a large hierarchical tree was generated that included all locations. We designated a number of clusters in each region to accommodate a wide variation of local densities. The coefficients obtained with the hierarchical clustering

[16] We orthogonalized the variables using command *orthog* in Stata 11 (StataCorp 2009).

method were similar in sign, significance, and magnitude to those in the previous tables.

Third, we repeated the analysis of self-citation ratios using both inventor and examiner citations. Recent research suggests that high levels of examiner citations are associated with low-quality patents (Alcácer and Gittelman 2006). Therefore, including these citations added a new set of observations—patents whose citations were 100% examiner imposed—that may represent inferior innovations. Results using citations from all sources were similar in magnitude and sign, but weaker in statistical significance. Finally, we reestimated the models with the top 5% of semiconductor firms (30 firms), still in 25 clusters, and obtained results similar in terms of sign, magnitude, and statistical significance.

6. Discussion

Although geographic collocation has obvious benefits for firm innovation, it can also have serious drawbacks. We explored how leading innovators can tap the rich resources in technology clusters while still appropriating value from their R&D investments. Our empirical findings suggest that internal linkages across multilocation firms play an important role in knowledge appropriation, even after controlling for knowledge sourcing opportunities. Specifically, by increasing control and interdependence across firm locations, internal linkages simultaneously facilitate knowledge internalization and reduce knowledge expropriation by nearby competitors. We also find that firms' strategic responses vary depending upon the neighbors with which they collocate. Firms tend to intensify internal linkages when neighboring firms share the same product market, but not when they merely overlap in the technological space.

We believe our study sheds light on some important aspects of location and innovation strategies. By studying the interaction between a firm's strategic internal organization and its location characteristics, we contribute to the appropriation literature that treats firms as abstract entities, and to the technology cluster literature that overlooks the complex internal organization of firms. We argue that a geographically dispersed organization, if managed properly, can provide a competitive advantage not only in knowledge sourcing but also in knowledge appropriation. By paying closer attention to the role of multilocation firms and their strategic organizations, we can refine our understanding of strategic dynamics in technology clusters. Our finding that firms respond to market competitors differently from technological competitors also adds a concrete location-specific context to the stream of literature on the interaction between technology and product-market strategies (Spence 1984, Kamien et al. 1992).

Our findings also have important implications for innovating firms and policy makers. For industry leaders making location decisions, our study shows that highly competitive technology clusters are not unduly risky, provided a firm actively manages its knowledge appropriation. The risk of exposing certain technologies to local competitors is low if those technologies are highly dependent on internal resources residing in another firm location. For policy makers eager to nurture local high-tech industries, our findings suggest that attracting leading innovators to a location is not, in itself, enough. Tax breaks and other incentives may influence where R&D is conducted, but not how R&D projects are actually organized. With local projects closely intertwined with the firms' global research agenda, the same R&D budget or R&D intensity may generate very different knowledge outflows to the local community.

This study does have limitations. For example, we focus on leading firms in one industry—an industry with high product modularity. Further analysis with more diverse contexts, including industries with low levels of modularity, will make our conclusions more generalizable. Also, by relying on patent data to capture innovation and knowledge flow, we leave out other forms of proprietary knowledge. Moreover, patents might be an imperfect source of information for cross-cluster teams (Bergek and Bruzelius 2010). Finally, there may be other mechanisms that allow industry leaders to alleviate expropriation concerns. For example, leading semiconductor firms are in a typical multimarket contact situation (Bernheim and Whinston 1990), in which competition may be attenuated. We cannot outline all of these potential mechanisms, but the fact that we observe stronger internal linkages in the presence of competitors—even with the possibilities of multimarket contact and other attenuating mechanisms—makes our estimates conservative.

We see several avenues for further inquiry. First, although the mechanisms we discussed are based on multilocation firms, the need to appropriate economic rents from proprietary innovation applies to any firm or organization. Some aspects of internalization strategies are more generally applicable, such as the separation of complementary components in an R&D project, but implementing these strategies gets much harder without access to the geographic separation used by multilocation firms. Further research on internal organization and knowledge appropriation could explore other types of strategic organization that are less location specific.

Second, the strategies we discussed are based on a well-established set of internal routines and organizational skills that facilitate the transfer and integration of geographically dispersed knowledge—essentially, the effect of internal linkages (an organization issue) on the assimilation of R&D knowledge (a technological issue) in the face of competition (a product market issue). However, because internal linkages take time to develop, their presence may not always be in sync with changes in technological or competitive environments. Moreover, not every firm can manage internal linkages with enough efficiency and cost effectiveness. Hence, it is important to understand how firm heterogeneity affects the applicability of these strategies.

Finally, our arguments excluded the possibility of interorganizational cooperation. If R&D is fragmented across the value chain and outsourced to specialized firms (Arora et al. 2001), knowledge flow across organizational boundaries is necessary and even desirable. Studies by Appleyard (1996) in the semiconductor industry and by Schrader (1991) in the specialty steel and mini-mill industry identified information sharing even among employees of direct competitors. These interactions limit firms' abilities to exercise strict knowledge internalization. More research could show precisely how firms protect and extract value from innovations developed within permeable, changing, and diffuse firm boundaries.

Acknowledgments

The authors thank the editor, associate editor, and three anonymous reviewers for their thoughtful comments on the previous version of this paper. The authors also benefited from valuable suggestions and comments made by participants in seminars at Stern School of Business, New York University; the Wharton School, University of Pennsylvania; Harvard Business School; Fuqua School of Business, Duke University; Carnegie Mellon University; Robert H. Smith School of Business, University of Maryland; Rutgers University; Rotman School of Management, University of Toronto; Olin Business School at Washington University; College of Management, Georgia Tech; Ross School of Business, University of Michigan; London Business School; Krannert School of Business, Purdue University; National University of Singapore; and Tuck School of Business at Dartmouth.

References

Agarwal, R., M. Ganco, R. H. Ziedonis. 2009. Reputations for toughness in patent enforcement: Implications for knowledge spillovers via inventor mobility. *Strategic Management J.* **30**(13) 1349–1374.

Alcácer, J. 2006. Location choices across the value chain: How activity and capability influence collocation. *Management Sci.* **52**(10) 1457–1471.

Alcácer, J., M. Gittelman. 2006. Patent citations as a measure of knowledge flows: The influence of examiner citations. *Rev. Econom. Statist.* **88**(4) 774–779.

Alcácer, J., M. Gittelman, B. Sampat. 2009. Applicant and examiner citations in U.S. patents: An overview and analysis. *Res. Policy* **38**(2) 415–427.

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

Appleyard, M. M. 1996. How does knowledge flow? Interfirm patterns in the semiconductor industry. *Strategic Management J.* **17**(Special Issue) 137–154.

Arora, A., A. Fosfuri, A. Gambardella. 2001. *Markets for Technology: The Economics of Innovation and Corporate Strategy*. MIT Press, Cambridge, MA.

Audretsch, D. B., M. B. Feldman. 1996. R&D spillovers and the geography of innovation and production. *Amer. Econom. Rev.* **86**(3) 630–640.

Bergek, A., M. Bruzelius. 2010. Are patents with multiple inventors from different countries a good indicator of international R&D collaboration? The case of ABB. *Res. Policy* **39**(10) 1321–1334.

Bernheim, B. D., M. Whinston. 1990. Multimarket contact and collusive behavior. *RAND J. Econom.* **21**(1) 1–26.

Cassiman, B., R. Veugelers. 2002. R&D cooperation and spillovers: Some empirical evidence from Belgium. *Amer. Econom. Rev.* **92**(4) 1169–1184.

Chung, W., J. Alcácer. 2002. Knowledge seeking and location choice of foreign direct investment in the United States. *Management Sci.* **48**(12) 1534–1554.

Cohen, W. M. 1995. Empirical studies of innovative activity, P. Stoneman, ed. *Handbook of the Economics of Innovations and Technological Change*. Wiley-Blackwell, Oxford, UK, 182–264.

Cohen, W. M., D. Levinthal. 1990. Absorptive capacity: A new perspective on learning and innovation. *Admin. Sci. Quart.* **35**(1) 128–152.

Cohen, W. M., R. R. Nelson, J. P. Walsh. 2000. Protecting their intellectual assets: Appropriability conditions and why U.S. manufacturing firms patent (or not). NBER Working Paper 7552, National Bureau of Economic Research, Cambridge, MA.

Delgado, M., M. Porter, S. Stern. 2010. Clusters, convergence and economic performance. Working paper, Harvard Business School, Boston.

Doz, Y., K. Wilson, S. Veldhoen, T. Goldbrunner, G. Altman. 2006. Innovation: Is global the way forward? Joint study by Booz Allen Hamilton and INSEAD.

Dushnitsky, G., J. M. Shaver. 2009. Limitations to interorganizational knowledge acquisition: The paradox of corporate venture capital. *Strategic Management J.* **30**(10) 1045–1064.

Edstrom, A., J. R. Galbraith. 1977. Transfer of managers as a coordination and control strategy in multinational organizations. *Admin. Sci. Quart.* **22**(2) 248–263.

Ellison, G., E. L. Glaeser. 1997. Geographic concentration in U.S. manufacturing industries: A dartboard approach. *J. Political Econom.* **105**(5) 889–927.

Fosfuri, A., M. S. Giarratana, A. Luzzi. 2008. The penguin has entered the building: The commercialization of open source software products. *Organ. Sci.* **19**(2) 292–305.

Frost, T. S., C. Zhou. 2005. R&D co-practice and "reverse" knowledge integration in multinational firms. *J. Internat. Bus. Stud.* **36**(6) 676–687.

Frost, T. S., J. Birkinshaw, P. C. Ensign. 2002. Centers of excellence in multinational firms. *Strategic Management J.* **23**(11) 997–1018.

Gilson, R. J. 1999. The legal infrastructure of high technology districts: Silicon Valley, Route 128, and covenants not to compete. *New York Univ. Law Rev.* **74**(3) 575–629.

Gittelman, M., B. Kogut. 2003. Does good science lead to valuable knowledge? Biotechnology firms and the evolutionary logic of citations patterns. *Management Sci.* **49**(4) 366–382.

Gupta, A. K., V. Govindarajan. 2000. Knowledge flows within multinational corporations. *Strategic Management J.* **21**(4) 473–496.

Hall, B. H., A. B. Jaffe, M. Trajtenberg. 2005. Market value and patent citations. *RAND J. Econom.* **36**(1) 16–38.

Hannah, D. R. 2005. Should I keep a secret? The effects of trade secret protection procedures on employees' obligations to protect trade secrets. *Organ. Sci.* **16**(1) 71–84.

Helfat, C. E. 1994. Firm-specificity in corporate applied R&D. *Organ. Sci.* **5**(2) 173–184.

Jaffe, A. B., M. Trajtenberg. 2005. *Patents, Citations, and Innovations: A Window on the Knowledge Economy*. MIT Press, Cambridge, MA.

Kamien, M., E. Muller, I. Zang. 1992. Research joint ventures and R&D cartels. *Amer. Econom. Rev.* **82**(5) 1293–1306.

Kogut, B., U. Zander. 1993. Knowledge of the firm and the evolutionary theory of the multinational corporation. *J. Internat. Bus. Stud.* **24**(4) 625–645.

Lahiri, N. 2010. Geographic distribution of R&D activity: How does it affect innovation quality. *Acad. Management J.* **53**(5) 1194–1209.

Lanjouw, J. O., M. Schankerman. 2004. Patent quality and research productivity: Measuring innovation with multiple indicators. *Econom. J.* **114**(495) 441–465.

Liebeskind, J. P. 1996. Knowledge, strategy, and the theory of the firm. *Strategic Management J.* **17**(Winter Special Issue) 93–107.

Marshall, A. 1920. *Principles of Economics*, revised 8th ed. MacMillan, London.

Nobel, R., J. Birkinshaw. 1998. Innovation in multinational corporations: Control and communication patterns in international R&D operations. *Strategic Management J.* **19**(5) 479–496.

Oxley, J., R. Sampson. 2004. The scope and governance of international R&D alliances. *Strategic Management J.* **25**(8–9) 723–749.

Pisano, G. P. 2006. Profiting from innovation and the intellectual property revolution. *Res. Policy* **35**(8) 1122–1130.

Porter, M. E. 1998. *On Competition*. Harvard Business School Press, Cambridge, MA.

Putnam, J. 1996. The value of international patent rights. Ph.D. thesis, Yale University, New Haven, CT.

Rajan, R., L. Zingales. 2001. The firm as a dedicated hierarchy: A theory of the origins and growth of firms. *Quart. J. Econom.* **116**(3) 805–851.

Saxenian, A. 1994. *Regional Advantage: Culture and Competition in Silicon Valley and Route 128*. Harvard University Press, Cambridge, MA.

Schrader, S. 1991. Informal technology transfer between firms: Cooperation through information trading. *Res. Policy* **20**(2) 153–170.

Shaver, J. M., F. Flyer. 2000. Agglomeration economies, firm heterogeneity, and foreign direct investment in the United States. *Strategic Management J.* **21**(12) 1175–1193.

Singh, J. 2008. Distributed R&D, cross-regional knowledge integration and quality of innovative output. *Res. Policy* **37**(1) 77–96.

Spence, M. 1984. Cost reduction, competition, and industry performance. *Econometrica* **52**(1) 101–121.

StataCorp. 2009. *Stata 11 Base Reference Manual*. Stata Press, College Station, TX.

Stuart, T. E. 2000. Interorganizational alliances and the performance of firms: A study of growth and innovation rates in a high-technology industry. *Strategic Management J.* **21**(8) 791–811.

Stuart, T. E., J. M. Podolny. 1996. Local search and the evolution of technological capabilities. *Strategic Management J.* **17**(Summer Special Issue) 21–38.

Teece, D. 1986. Profiting from technological innovation: Implications for integration, collaboration, licensing, and public policy. *Res. Policy* **15**(6) 285–305.

Trajtenberg, M., R. Henderson, A. Jaffe. 1997. University versus corporate patents: A window on the basicness of invention. *Econom. Innovation New Tech.* **5**(1) 19–50.

Zhao, M. 2006. Conducting R&D in countries with weak intellectual property rights protection. *Management Sci.* **52**(8) 1185–1199.

Ziedonis, R. H. 2004. Don't fence me in: Fragmented markets for technology and the patent acquisition strategies of firms. *Management Sci.* **50**(6) 804–820.