Strategic Management Journal
Strat. Mgmt. J. (2015)
Published online EarlyView in Wiley Online Library (wileyonlinelibrary.com) DOI: 10.1002/smj.2363
Received 16 September 2013; Final revision received 9 September 2014

INNOVATION ECOSYSTEMS AND THE PACE OF SUBSTITUTION: RE-EXAMINING TECHNOLOGY S-CURVES

RON ADNER[1] and RAHUL KAPOOR[2]*
[1] *Strategy and Management, Tuck School of Business, Dartmouth College, Hanover, New Hampshire, U.S.A.*
[2] *Management Department, The Wharton School, University of Pennsylvania, Philadelphia, Pennsylvania, U.S.A.*

Why do some new technologies emerge and quickly supplant incumbent technologies while others take years or decades to take off? We explore this question by presenting a framework that considers both the focal competing technologies as well as the ecosystems in which they are embedded. Within our framework, each episode of technology transition is characterized by the ecosystem emergence challenge that confronts the new technology and the ecosystem extension opportunity that is available to the old technology. We identify four qualitatively distinct regimes with clear predictions for the pace of substitution. Evidence from 10 episodes of technology transitions in the semiconductor lithography equipment industry from 1972 to 2009 offers strong support for our framework. We discuss the implication of our approach for firm strategy.

INTRODUCTION

While the waves of creative destruction regularly crash on the shores of markets, the pace of substitution varies markedly across different contexts and episodes (e.g., Anderson and Tushman, 1990). Why do some new technologies immediately supplant incumbent technologies while others take decades to take off? Despite the centrality of the phenomenon in the literature, the pace of technology substitution has been under-explored.

At the level of technologies, the strategy literature has focused on whether new technologies will rise to dominate old technologies (e.g., Adner, 2002; Christensen, 1997; Foster, 1986) but largely ignored the question of when dominance will be achieved. Conversely, the diffusion of innovation literature has focused on the rate of adoption of new technologies (e.g., Hall, 2004; Rogers, 2003), but has taken a static view of the diffusing innovations (i.e., overlooking the continued evolution of both the new and old technologies). At the level of firms, the strategy literature has focused on how firms' capabilities and experience shape their entry decisions and performance in new technologies (e.g., Franco *et al.*, 2009; King and Tucci, 2002; Mitchell, 1989, 1991) but has tended to ignore the pace with which the new technology substitutes the old. Hence, while the emphasis has been on firm-level competition in a new technology, the technology-level competition between the new and old has been overlooked. Absent such a technology-level consideration, the literature is handicapped in offering guidance regarding how firms should manage the transition from old to new technologies, and in clarifying

Keywords: innovation; substitution; ecosystems; technology evolution; technology competition
*Correspondence to: Rahul Kapoor, The Wharton School, University of Pennsylvania, Philadelphia, PA 19104, U.S.A. E-mail: kapoorr@wharton.upenn.edu

when persisting with the old technology may represent a more viable strategy than aggressively pursuing the new technology.

In this paper we argue that understanding the pace of substitution requires joint consideration of the evolution of both the new and the old technologies. In turn, understanding this evolution requires an examination of the interdependencies in the broader ecosystem of components and complements in which the focal technologies are embedded (Adner, 2006, 2012; Adner and Kapoor, 2010; Christensen and Rosenbloom, 1995; Hughes, 1983; Iansiti and Levien, 2004; Moore, 1993; Rosenberg, 1976, 1982). Developing such an understanding has important implications for how firms pursue new and old technology opportunities (e.g., Christensen, 1997; Foster, 1986), and how firms manage interdependencies within their ecosystems as technologies evolve (e.g., Ethiraj, 2007; Kapoor and McGrath, 2014).

We present a structured framework to analyze the pace of technology substitution that considers the differential impact of ecosystem on the new and old technologies. On the one hand, bottlenecks that arise as other ecosystem elements struggle to emerge can act to constrain the competitiveness of the rising new technology (Hughes, 1983; Rosenberg, 1976, 1982)–they create ecosystem emergence challenges. On the other hand, advances in other ecosystem elements that arise after the old technology has matured can act to enhance the competitiveness of the old technology (Harley, 1971; Tripsas, 2008; Utterback, 1994)–they create ecosystem extension opportunities. Our joint consideration of ecosystem emergence challenges for the new technology and ecosystem extension opportunities for the old technology allows us to identify four qualitatively distinct substitution regimes that, in turn, give rise to a testable prediction regarding the pace of substitution across these regimes.

We test our theory in the context of the semiconductor lithography equipment industry from 1972 to 2009, a period during which 10 new technology generations were introduced into the market. These transitions offer an interesting puzzle: across the 10 generations there was remarkable absence of variance across the key factors that the literature has identified as driving substitution. In each case, the new technology generation was introduced into the marketplace on a commercial basis (i.e., it had overcome its development challenges, e.g., Henderson and Clark, 1990); at the time of market introduction, the new generation offered unambiguously superior performance relative to the predecessor generation on both an absolute performance and price-adjusted-performance basis (e.g., Foster, 1986); customers and the semiconductor manufacturing firms were well informed about the availability of the new generation (e.g., Bass, 2004; Rogers, 2003); and customers were eager to adopt higher performance technologies (e.g., Christensen, 1997). Moreover, each of the new generations preserved the core competences (Tushman and Anderson, 1986) and complementary assets (Tripsas, 1997) of the lithography equipment firms. Yet despite this constancy in conditions, the pace of substitution varied dramatically across the 10 technology generations, ranging from cases of rapid substitution (market dominance achieved after a single year), to slow substitution (dominance achieved after 10 years), to non-substitution (dominance never achieved).

We draw on a unique array of qualitative and quantitative data collected during two years of fieldwork to explore this puzzle. We find that our approach of explicitly considering the interaction between the old and new technology ecosystems explains significant variance in the observed pace of substitution in semiconductor lithography. Our theory and our observations are at the level of technology. However, through our fieldwork, we are also able to shed new light on how heterogeneity in firms' choices and investments shaped the path of technology transitions. We found that at times, the pace of substitution was slowed due to "last gasp" efforts by some firms to extend and maximize the value that they could capture from the old technology while other firms shifted aggressively to the new technology. In other instances, the discovery of solutions to the emergence challenges in the new technology allowed some firms to benefit from spillovers that extended the performance of the old technology. Finally, we observed a third mechanism that has not been previously characterized in the literature: homogeneous actions by heterogeneous firms as actors across the ecosystem–competitors, suppliers, complementors, users–engaged in a "last resort" effort to extend the old technology when confronted by a collective inability to overcome the emergence challenges of the new technology in the expected timeframe.

Our study illustrates the importance of considering ecosystem dynamics of the new technology's emergence and the old technology's extension to

DOI: 10.1002/smj

explain the pace of substitution. In doing so, it informs the literatures on technology, strategy and innovation diffusion by showing how the interactions between old and new technology ecosystems shape the context for both competition and market adoption. Our perspective is not one of technological determinism but rather one of understanding the context and constraints that impact firm choices and outcomes during technology transitions. In the concluding section, we consider the implications of our approach and framework for managers and policy makers.

TECHNOLOGY SUBSTITUTION

Explaining the dynamics of substitution is an important goal of the technology strategy literature. S-curves have become the canonical representations of both the technology life cycle and of the competition between technologies (e.g., Christensen, 1997; Foster, 1986; Utterback, 1994). The S-curve approach holds that the magnitude of performance improvement in a given technology for a fixed unit of effort or time is relatively low during the early development stages. As the technology is better understood, the rate of progress increases until the stage of maturity, at which point the technology approaches its limits and the performance impact of additional effort is subject to decreasing returns. In the context of competing technologies, Foster's influential work posited that the substitution threat from the new technology becomes salient once it has superior performance–that is, whether a new technology will dominate the market depends on whether the new S-curve crosses the old.

The S-curve characterization of technology evolution has been critiqued and refined along a number of dimensions. Sood and Tellis (2005) show that the profile of technology improvement over time can vary markedly from the smooth S shape. Christensen (1997) showed that substitution can take place even when the new technology is inferior to the old, arguing that if users are over-served along the main performance dimension they may switch to the new technology if it offers superior performance on a different dimension. Adner (2002) found that such disruptive dynamics can be explained in terms of price and cost asymmetries between producers of the old and new technologies. Other studies have shown how new technologies can be incubated in distant markets, emerging as substitute threats when they reach a sufficient performance level (Adner and Levinthal, 2001; Levinthal, 1998) or when triggered by a discontinuous shift in the preferences of mainstream consumers (Tripsas, 2008).

Notice, however, that studies in this vein implicitly frame substitution as an event that is governed by the rise of the new technology. They pay the bulk of their attention to understanding whether the new technology arises in a market domain that is nearer or farther from the current customer base; whether it draws on knowledge domains that are nearer or farther from the current competence base; whether it favors established firms or new entrants. Throughout, the focus is on whether substitution will occur, leaving the question of how fast substitution will unfold unaddressed.

In contrast, the diffusion of innovation literature explicitly considers the factors that determine the rate of market adoption. This literature takes the availability of a given innovation as its starting point and then considers how characteristics of the new technology interact with its users and their social context to impact its rate of adoption. Rogers' comprehensive treatment (2003) identifies five canonical attributes of innovations that determine the rate of their adoption (relative advantage; compatibility; complexity; trialability; observability) as well as contextual factors such as the nature of communication channels, social systems, and promotion efforts. The diffusion literature also considers factors such as network effects (Stremersch *et al.,* 2007), standards (Dranove and Gandal, 2003), and information contagion (Iyengar, Van den Bulte, and Valente, 2011).

However, while the diffusion literature has been articulate about the factors that impact adoption dynamics, it has tended to take a static view of the technology itself. As Hall notes in her review, the diffusion literature "implicitly assumes that neither the new innovation nor the technology it replaces changes during the diffusion process." (2004: 5). New technology improvements, however, are at the heart of the technology strategy literature (e.g., Christensen, 1997; Foster, 1986). And although the canonical representation posits a flat-lining of the old technology's performance trajectory, even mature technologies can be reinvigorated. Cases of old technology extension have been well documented in settings ranging from sailing ships (Harley, 1971) to pond ice harvesting (Utterback, 1994) to typesetters (Tripsas, 2008) to carburetors

DOI: 10.1002/smj

(Snow, 2008) to mobile telephony (Ansari and Garud, 2009). Thus, while the diffusion literature focuses on the question of how fast the new technology gains market share, it does so without consideration of the evolution of the new and the old technologies.

Linking substitution and adoption through systems

The technology strategy and the diffusion literatures approach the question of substitution from different directions. Whereas the former focuses on the supply-side concerns of firms developing higher performance technologies, the latter focuses on the demand-side concerns of users adopting higher value technologies. A deeper understanding of substitution dynamics requires linking technology evolution to technology adoption. The systems approach to technology offers just such a link.[1]

When a technology is consumed as part of a system, the performance that matters to the consumers' assessment of value–the realized performance–is not the performance of the focal technology on its own, but is rather a function of its interaction with the other elements of the system. Here we distinguish between component elements that are integrated by the focal firm into its offer and complement elements that are integrated by the user in conjunction with the focal firm's offer (Adner and Kapoor, 2010).

The system can have two distinct effects on the realized performance of a technology. First, the realized performance of a technology can be hindered by technical bottlenecks within the system. As Rosenberg (1976) notes, "… a given invention, however promising, often cannot fulfill anything like its potential unless *other* [emphasis in original] inventions are made relaxing or bypassing constraints which would otherwise hamper its diffusion and expansion." (p. 201). His account of the American machine tool industry points to numerous episodes of "technical imbalance" in which improvements in new generations of machine tools could not be realized by users because of bottlenecks in the development of complements such as drills and cutting materials. It was only after the emergence challenges of these complements were resolved that the new generations of tools could finally deliver performance *that matched their potential*, and market adoption could begin in earnest. Similarly, Hughes (1983) rich description of the evolution of the electrical power system highlights the impact of technological interdependencies and performance bottlenecks on adoption.

Second, the realized performance of a technology can be enhanced when improvements in ecosystem elements offset maturity in the focal technology. Numerous accounts have shown that underlying a technology's advance are not only efforts by producers of the focal technology, but also systemic efforts by component and complement providers from a range of interdependent industries. For example, Constant's (1980) study of the rise of aircraft piston-engine technology pointed to the critical role played by improvements in engine components by suppliers such as General Electric, as well as improvements in the complements of fuels and lubricants by oil companies such as Royal Dutch Shell. Similarly, Henderson's pioneering (1995) study identified the key role of improvements in components and complements in extending the performance trajectory of optical lithography.

The pace of technology substitution: emergence challenges vs. extension opportunities

Recognizing the role of the system in constraining and extending the realized performance of the technology allows us to recast the question of substitution not as a competition between a new technology and an old technology, but rather as a competition between a new technology's ecosystem and an old technology's ecosystem.

[1] Our systems-based approach to technology transitions is closest to Adner and Kapoor (2010) and Ansari and Garud (2009). Adner and Kapoor (2010) examined the role of ecosystem challenges in affecting the extent of (firm-level) early mover advantage within a new technology generation. We complement that study by considering the effects of ecosystem dynamics, of both the old and new technologies, in governing the pace of (technology-level) transitions. Ansari and Garud (2009) examined the influence of the socio-technical system on the transition between second generation (2G) and third generation (3G) mobile communication networks. Whereas their study presents conjectures about the direct effects of emergence and extension in the context of a single technology transition, we take the interaction between these forces as our central construct and show why it is critical to examine them jointly. We complement their work by developing a framework that considers the degree of emergence challenges and extension opportunities across the new and old technologies, and identify how and why their interaction can give rise to qualitatively distinct substitution regimes. Moreover, we offer a systematic examination of pace of substitution over the course of 10 generational transitions that helps to account for several alternative explanations.

DOI: 10.1002/smj

Ecosystem Extension Opportunity (Old Technology)

Ecosystem Emergence Challenge (New Technology)		Low	High
	Low	*Quadrant 1* Baseline pace of substitution	*Quadrant 2* Intermediate pace of substitution
	High	*Quadrant 3* Intermediate pace of substitution	*Quadrant 4* Slowest pace of substitution

Figure 1. A framework for analyzing technology substitution

Accordingly, we must consider not only the performance of the new technology, but also the extent to which technology bottlenecks elsewhere in the system constrain the new technology's realized performance from matching its potential performance–the ecosystem emergence challenge.[2] Similarly, for the old technology, we must expand our consideration to include the extent to which improvements elsewhere in the system enhance the realized performance attainable with the old technology system–the ecosystem extension opportunity. Figure 1 presents our organizing framework and characterizes the joint effects of these two distinct constructs on the pace of substitution. Each quadrant represents the conditions that characterize the extent of new technology's ecosystem emergence challenge and old technology's ecosystem extension opportunity. The specific balance of these two forces gives rise to four different substitution regimes.

[2] Our characterization of ecosystem emergence challenges as rooted in technical bottlenecks is qualitatively distinct but fully complementary with insights regarding the impact of standards and direct and indirect network effects (e.g., Katz and Shapiro, 1985; Farrell and Saloner, 1986; Dranove and Gandal, 2003) on adoption and substitution. The distinction lies in the difference between a partner's *challenge* to provide a complement and a partner's *choice* to provide the complement. In addition, standards wars, direct network externalities, and indirect network externalities focus on the impact of the *number* of adopters (installed base) and complementors who *choose* to support a given technology. In contrast, the question of emergence challenges focuses on the very possibility of participating -- the extent to which technical bottlenecks prevent eager partners from delivering a working complement that would allow the new technology to translate its superior technical performance to superior value creation. Here, the constraint to the availability of the complement is *not* rooted in issues of adoption incentive and critical mass, but rather in the complementors' own capabilities.

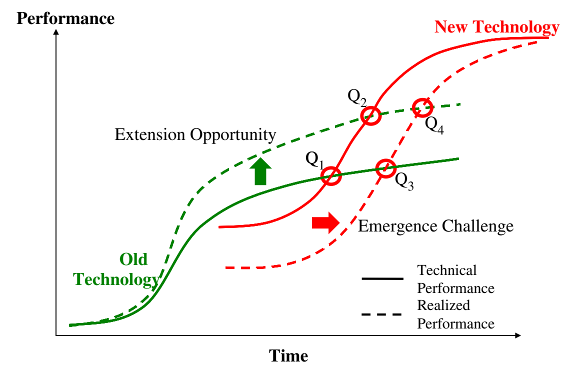


Figure 2. Technology competition between an old technology with ecosystem extension opportunity and a new technology with ecosystem emergence challenge

Figure 2 offers a graphical intuition for the implication of these system-level technology dynamics for the relative pace of substitution. The baseline case (the two solid S-curves, corresponding to Quadrant 1 in the framework) is the one in which neither technology's performance is impacted by interaction with ecosystem elements, such that technical performance fully determines the realized performance: the new technology does not face ecosystem emergence challenges and the old technology does not benefit from ecosystem extension opportunities. Q1 marks the point at which the performance of the new technology exceeds that of the old; and therefore, the time at which we can expect adoption to accelerate with market dominance to follow.[3]

[3] The figure illustrates the classic pattern of competing S-curves in which the performance of new technology is initially inferior to that of the old technology. This is clearly not always the case (Sood and Tellis, 2005). Indeed across the 10 generational transitions

DOI: 10.1002/smj

When the extension-emergence balance shifts, however, we can expect very different dynamics. In diametric opposition to the prior case, is the case when the ecosystem emergence challenge for the new technology is high and the ecosystem extension opportunity for the old technology is high as well (the two dashed S-curves, corresponding to Quadrant 4 in the framework). Here, despite the promise of the new technology, its realized performance is constrained by bottlenecks elsewhere in the ecosystem. Moreover, the old technology stands to benefit from improvement in its own components and complements that enhance its realized performance, even if its technical performance is stagnant. Q4 marks the point at which the realized performance of the new technology exceeds that of the old; thus, the simultaneous presence of emergence challenge and extension opportunity shifts out the time at which we can expect adoption to accelerate.

When the emergence challenge for the new technology is low and the extension opportunity for the old is high (Quadrant 2), competition between the old and new technology will be robust. The new technology will make inroads into the market, but improvements in the incumbent technology's ecosystem allow it to compete effectively. We expect a prolonged period of coexistence before substitution takes place (e.g., Furr and Snow, 2014; McGahan, 2004). Although high extension opportunity is unlikely to reverse the rise of the new technology, it will materially delay the new technology's market dominance and slow the rate of substitution. Thus Q2, which marks the transition point, occurs later than Q1 but earlier than Q4.

In contrast, when the emergence challenge is high for the new technology and the extension opportunity is low for the old technology (Quadrant 3), substitution will be delayed until emergence challenges are resolved. We therefore expect Q3 to occur later than Q1 but earlier than Q4. We thus expect the pace of substitution in both Quadrants 2 and 3 to fall between Quadrants 1 and 4.

The joint consideration of the system-level technology dynamics for the competing old and new technologies allows us offer the following prediction:

Hypothesis: The relative pace at which the new technology will substitute the old technology will depend on the joint levels of ecosystem emergence challenge for the new technology and the ecosystem extension opportunity for the old technology: fastest when both are low, slowest when both are high, and intermediate in the mixed case where one is high and one is low (i.e., Q1 < Q2, Q3 < Q4).

Although we are agnostic about the specific ordering for Quadrants 2 and 3, we do expect the underlying technology dynamics within these quadrants to be qualitatively different: In the case of Quadrant 2, substitution is slowed due to continued improvements in the old technology ecosystem. Here we would expect a gradual shift in market shares from the old to the new technology. In contrast, in the case of Quadrant 3, the old technology's dominance is a function not of improvements in its ecosystem, but rather of the new technology's ecosystem emergence challenge. Substitution will be delayed until emergence challenges are resolved, but once they are resolved the transition in market share will be rapid. Here, an analysis of market share will likely show that the old technology maintains high market share, but that market growth has stalled. Because rapid market share inversion is to be expected once the new technology fulfills its performance potential, dominance by the incumbent technology is fragile, in the sense that it is maintained not by its own progress, but by the new technology's emergence challenge. Thus, while our formal hypothesis focuses on the pace of substitution, we expect that our logic can inform the market-share profile of substitution as well.

We note that much of the literature on technology substitution has tended to focus on the performance of competing focal technologies without considering the ecosystem emergence challenges of new technologies and the ecosystem extension opportunities of old technologies (e.g., Anderson and Tushman, 1990; Bass, 2004; Christensen, 1997; Fisher and Pry, 1972; Foster, 1986; Sood and Tellis, 2005). Thus, Quadrants 2, 3 and 4 fall outside the scope of mainstream analyses. The

that we study in the semiconductor lithography equipment industry every new technology is introduced with an initial performance level that is substantially superior to that of the old technology. Regardless of initial technology positions, however, the implications of emergence challenges and extension opportunities remain the same, hindering and enhancing the new and old technologies respectively.

sub-strand of the literature that studies systems of technology has tended to focus only on the emergence challenges of the new technology (Hughes, 1983; Rosenberg, 1976). Similarly, the literature on competing systems (e.g., Cusumano *et al.*, 1992; Katz and Shapiro, 1985) explores the relative emergence challenges faced by competing new technologies, but overlooks the role of the old technology ecosystem in shaping competition between old and new technologies. Thus, it neglects Quadrants 2 and 4 of our framework. Conversely, the sub-strand that has focused on the improvement in the old technology (e.g., Henderson, 1995; Utterback, 1994) has tended to overlook the role of new technology's emergence challenge in shaping technology competition (i.e., Quadrants 3 and 4). It is our explicit consideration of the *joint* effects of the old and new technology ecosystems that allows us to characterize four distinct substitution regimes in a single framework giving rise to testable predictions and distinct strategic implications.

METHODOLOGY AND DATA

We apply our framework to analyze technology transitions in the semiconductor lithography equipment industry from 1972 to 2009. The industry offers a particularly rich setting in which to explore the interaction between technology ecosystems and the dynamics of substitution. It is a context of robust technological change. Between 1972 and 2009, 10 new generations of lithography equipment technology were commercialized, each enabling semiconductor manufacturers, the customers, to achieve exponential improvements in chip performance at lower marginal production cost along the trajectory referred to as Moore's law. While lithography equipment firms played an important role in pushing the technology envelope forward, progress within each of the technology generations was critically shaped by components and complements in the ecosystem (Adner and Kapoor, 2010; Henderson, 1995; Iansiti, 1998). Simultaneously, the context was characterized by continuity in terms of user preferences (Christensen, 1997; Tripsas, 2008), nature of equipment manufacturers' core competences (Anderson and Tushman, 1990) and complementary assets (Teece, 1986), and low network effects (Katz and Shapiro, 1985).

The research setting also provides natural controls for key factors considered in the innovation of diffusion literature that affect the pace of substitution (Rogers, 2003). Throughout the industry's history, the new technology generations were characterized by a high degree of observability and trialability by the users. This is because lithography equipment firms aggressively promoted their new generation "tools" through trade shows, industry conferences as well as through formal and informal interactions with semiconductor manufacturers well before the new generation was commercialized. Similarly, customers saw the new technology as key to their own competitive advantage and therefore dedicated personnel within their organizations to actively scan and evaluate new technology developments. Moreover, given the high scale of investment and the criticality of the lithography tool to their manufacturing process, semiconductor manufacturers conducted extensive experimentation prior to purchasing and implementing the new technology generations into their processes. In our analysis, we account for the additional diffusion factors of relative advantage, compatibility and complexity of the new technology generation.

The nature of this study required us to develop a thorough understanding of the lithography technology, its historical evolution as well as the nature of emergence challenges and extension opportunities within the ecosystems of the different technology generations. We interviewed over 30 industry experts, most of whom have been associated with the industry for more than 20 years. The interviews were semi-structured and lasted two hours on average. The interviewees came from a variety of roles within the semiconductor lithography ecosystem: lithography equipment manufacturers, their customers, their complementors, their suppliers, research consortia, industry associations, and consultants.[4] This allowed us to understand the emergence challenges and extension opportunities across the ecosystem for each of the 10 technology transitions and to triangulate perspectives across different roles.

For the quantitative analysis, we obtained data on detailed product specifications and sales for each of the technology generations from VLSI Research, a prominent industry consulting firm.

[4] Twenty-seven percent of the interviewees were affiliated with lithography equipment manufacturers, 24 percent with customers, 17 percent with complementors, 12 percent with suppliers and the remainder with SEMATECH, Industry associations and consultants.

DOI: 10.1002/smj

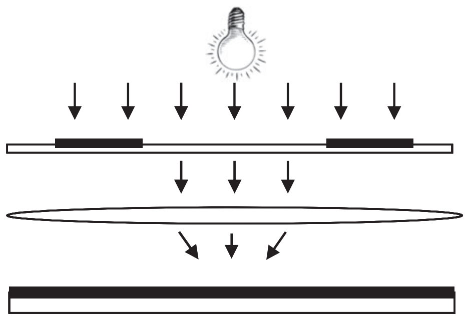


Figure 3. Basic principle of the semiconductor lithography process

VLSI Research also provided us with historical data on prices of lithography equipment and capital expenditures by semiconductor manufacturers that we used to rule out some important alternative explanations. For two of the earliest generations, we supplemented the data from VLSI Research with product specifications reported in the Henderson Photolithography FIVE Data (Helfat and Klepper, 2007; Henderson, 1993). Through our fieldwork, we also developed a list of keywords that we used to codify technical articles published between 1961 and 2010 in *Solid State Technology,* the leading industry trade journal, and we used this information to develop an objective measure of ecosystem emergence challenge for each of the technology generations.

SEMICONDUCTOR LITHOGRAPHY TECHNOLOGY

Semiconductor lithography is the process by which a circuit design is imprinted on a semiconductor substrate called a wafer. The basic principle of lithography is illustrated in Figure 3. After the design of an integrated circuit (IC) is finalized (i.e., the wiring, the gates and the junctions), the circuit blueprint is transferred to a "mask." The lithography process takes place when beams of energy originating from an "energy source" are directed onto the mask. The pattern on the mask allows a portion of the energy beams to pass through, with or without an optical "lens" system, onto the wafer. The wafer is coated with an energy sensitive "resist." The resist undergoes a chemical reaction wherever the mask has allowed the energy to pass through. This chemical reaction changes the structure of the resist and allows its selective removal from the wafer. Another chemical process is then initiated in which the exposed parts of the wafer are etched. Finally, the remaining resist is removed, creating a final circuit that replicates the initial design.

The process is carried out using lithography alignment equipment referred to in the industry as the lithography tool. The energy source and the lens are the two key components that are integrated by equipment manufacturers into the lithography tool. A semiconductor manufacturer may use scores of lithography tools in a single production line. With modern tools costing over twenty million dollars each, investments in lithography equipment represent a substantial portion of the cost of a fabrication facility. To maintain competitiveness, semiconductor manufacturers continuously reinvest in their facilities and look to new tool generations to allow them to offer products with higher performance at lower marginal cost. The mask and the resist are the two main complements that semiconductor manufacturers must integrate with the lithography tool to carry out the lithography process. Hence, the key elements of the lithography technology ecosystem are the focal technology (the lithography tool), the energy source and the lens as the key components, and the mask and the resist as the key complements (Figure 4).[5]

The main measure of performance in semiconductor lithography is resolution, the smallest feature size that can be "printed" on the semiconductor wafer. Resolution determines the extent of miniaturization that can be achieved by the semiconductor manufacturer. Smaller feature size allows the semiconductor manufacturers to pack more circuits onto a chip and more chips onto a wafer. What this means is that for both the lithography

[5] Three of the earliest technology generations transmitted the energy directly from the source to the wafer without using a lens. These were the Contact, Proximity, and X-ray generations.

DOI: 10.1002/smj

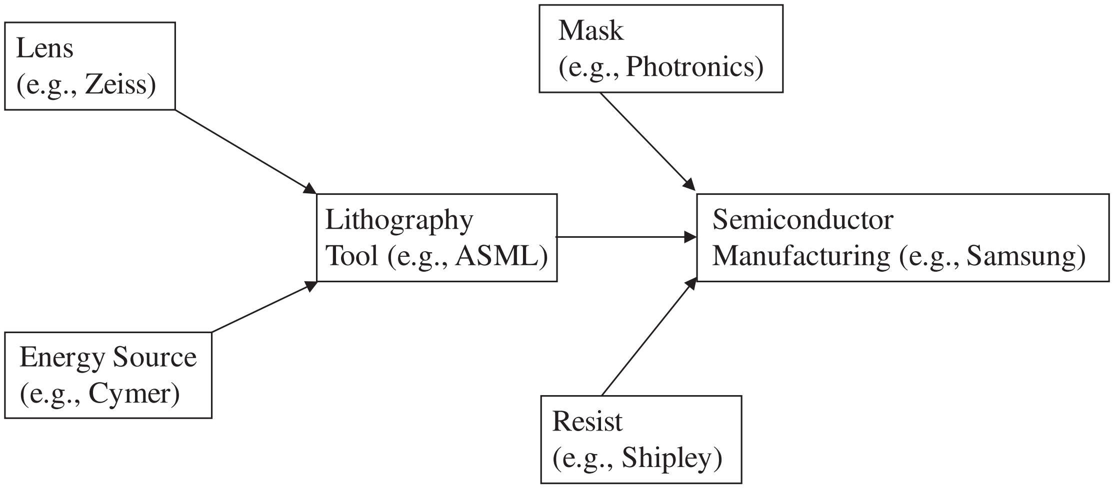


Figure 4. Semiconductor lithography ecosystem. Firms mentioned in parenthesis are only representative of the firms participating in the lithography ecosystem

tool manufacturers and their customers, the semiconductor manufacturers, improving resolution is the key to maintaining competitive advantage and is therefore a top priority. Although other dimensions of performance matter (e.g., throughput, service reliability), resolution is by far the dominant consideration.[6]

TECHNOLOGY SUBSTITUTION IN THE SEMICONDUCTOR LITHOGRAPHY EQUIPMENT INDUSTRY

The semiconductor lithography equipment industry distinguishes among generations according to the changes in the tool architecture and in the wavelength of light. From 1972 to 2009, 10 distinct generations of lithography equipment have been introduced.[7] We provide a detailed description of each of the generations in the Appendix I including information on the nature of ecosystem emergence challenge and the years it took to achieve a dominant market share. Judged according to the established criteria that have been identified in the literature, the conditions of their launches were nearly identical: At the time of first commercial sale, the resolution performance of each new technology generation was superior to the performance offered by the old technology. Each of the generations were also sustaining (Christensen, 1997) in that they were developed to meet the needs of firms' existing customers who were demanding improved resolution so as to manufacture integrated circuits (ICs) with higher performance and lower cost. Finally, the composition of key customer segments, the principle dimensions of performance, as well as the modes through which producers interacted with customers (Abernathy and Clark, 1985) were all consistent throughout the industry's history.

Given these structural similarities, conventional wisdom would yield a prediction that the technology generations should follow similar substitution patterns. This, however, is not the case. Indeed, there is significant variance in the time it took for the new generation to achieve a dominant market share. There are cases of fast substitution in which the new technology takes off almost immediately upon its introduction and achieves the dominant market share. These include the Proximity, Projection, G-line and DUV 193-immersion generations. However, there are also cases of slow transitions in which

[6] In File S1 that is available as an online supplement, we present a hedonic regression analysis of lithography tools and find that resolution on its own explains 81 percent of observed variance in price.

[7] This definition of lithography generation is consistent with the industry norm as well as Henderson (1993) and Henderson and Clark (1990). It offers a finer grained level of analysis than the general category of "optical lithography" used in Henderson (1995), which would group eight of the 10 generations examined here (all but E-beam and X-ray) under the single heading. As discussed below, focusing on the transitions between individual generations offers a more nuanced view of substitution. It also leads us to characterize the failure of X-ray lithography, discussed in Henderson (1995) as driven in large part by its own emergence challenges rather than purely by the successful extension of optical lithography.

DOI: 10.1002/smj

the new technology takes much longer to achieve market dominance. These include the I-line, DUV 248 and DUV 193 generations. Finally, there were cases of non-transitions in which despite significant development efforts by the industry over many years, the new technology failed to substitute the old technology. These include the E-beam, X-ray and DUV 157 generations.[8]

Emergence challenge in the new technology generation

Each of the new generations presented significant development challenges for the lithography equipment manufacturers. These include challenges in designing the new tool architecture as well as in integrating the different components into the architecture. At the time of first commercial sale, each new generation offered superior technical performance than did the dominant generation. However, these generations differed significantly in the extent to which innovation in complements was required in order for the new generation to achieve its commercialization potential; that is, for its realized performance to match its technical performance for the mass market. As detailed in the Appendix I, 5 of the 10 generations faced high emergence challenges, requiring major complement innovations in the mask and resist. These were the E-beam, X-ray, DUV 248, DUV 193 and DUV 157 generations. The other five generations–Proximity, Projection, G-line, I-line and DUV 193-immersion–faced much lower emergence challenges as they were able to reuse the mask and resist technologies that had been developed for the old technology generations.

For an objective measure of ecosystem emergence challenge we follow Adner and Kapoor (2010) and use a count of the number of technical articles published in *Solid State Technology*, the leading industry journal. We counted articles that reported on technical challenges regarding the mask and resist for each of the new technology generations.[9] A representative discussion of emergence challenge is Hibbs, Kunz, and Rothschild (1995: 69):

> "The most difficult challenge in bringing 193-nm lithography to full manufacturing use is the development of a robust photoresist process. The resins that are typically used for i-line and 248-nm photoresists (novolac, polyhydroxystyrene, etc.) are far too opaque at 193 nm to be used in single-layer resists at that wavelength. Other materials, such as acrylate polymers, have been identified as suitable for 193-nm resists. Using these materials, prototype resists with good imaging properties have been developed in a collaborative effort between IBM/Almaden and Lincoln Laboratory. However, the dry etch resistance of these prototype resists is relatively poor, and modifications of the formulation are being developed to provide higher etch resistance."

As this excerpt, which was published a year in advance of the first commercial sale of a DUV 193 tool, shows that the resist challenge was a genuine technology bottleneck, rather than a problem of insufficient incentives for complementors or a lack of a critical installed base of users.

The count of articles for each new generation is plotted in Figure 5. The average number of articles across the generations is 4.6. We categorized emergence challenges as "High" and "Low" according to whether a given generation was above or below this average. The characterization of ecosystem

[8] Other technology generations such as Ion Beam and Electron Projection were also pursued in the 1970s but did not move beyond R&D to the commercialization stage and, hence, are not part of our study.

[9] We used a number of keywords to identify lithography related articles that were published between 1961 and 2010. We categorized articles as lithography related if any of the following key words appeared in the title: lithography, microphotograph, mask, photomask, resist, laser, UV, DUV, Deep UV, optical, lens, stepper, aligner, mercury, illuminator, exposure, printer, and the names of the different generations. We then used the article titles to create a match between the technology generation and whether the article referred to technical problems in mask and resist. If insufficient information was available in the title, we read the abstract and the conclusion to ascertain if the article addressed the innovation challenges in the ecosystem for a given generation. A small subset of articles discussed the mask and resist challenges for multiple generations, and for these articles we read the relevant sections for each generation in order to create a match. Our industry sources confirmed that a count of published articles is a good proxy for the level of technology challenges that surrounded the development of different lithography generations. Since our primary concern was to assess challenges during the emergence of the new technology generation, we excluded articles that were published five years after the generation was commercialized. As a robustness check, we also used a 10 year window and found very similar patterns as those reported here.

DOI: 10.1002/smj

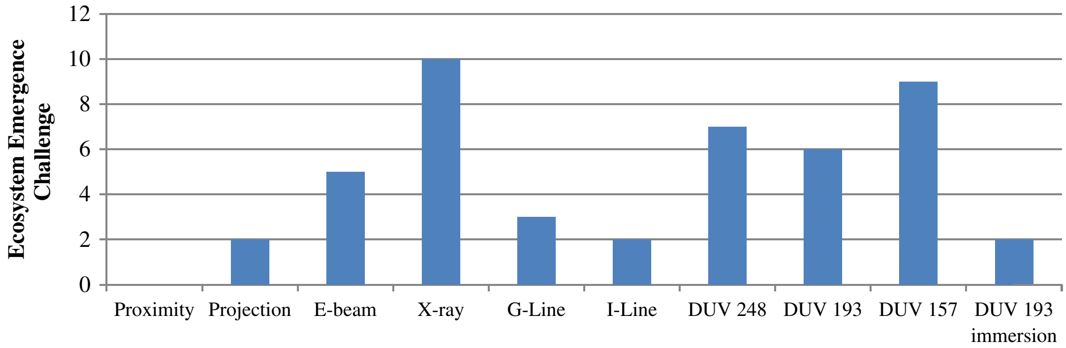


Figure 5. Magnitude of ecosystem emergence challenge for the different technology generations as measured by *Solid State Technology* article count

emergence challenge based on the article count exhibited very high consistency with the qualitative account of ecosystem emergence challenges that we document in the Appendix I.

Extension opportunity in the old technology generation

The different lithography technologies faced very different opportunities for extending their resolution performance at the time of the introduction of the new technology generation. The resolution of a technology generation is based on the theoretical relationship known as the Rayleigh criterion:

$$Resolution = k_1 \times \lambda / NA,$$

where λ is the wavelength of energy being transmitted by the source, NA is a numerical aperture of the lens that is a function of the tool architecture and the size of the lens, and k_1 is a process-specific constant that is a function of the mask and the resist complements. Within a given generation, λ is fixed. Therefore, the primary means of extending the performance of a technology generation are to improve the mask and the resist complements so as to reduce k1 or to improve the lens component so as to increase the NA. The mode of improvement in the mask and the resist complements are limited by the properties of the materials and the associated manufacturing techniques. The mode of improvement in the lens component is limited by the complexity of the lens design and manufacturing, as well as the reduction in the allowable margin of error for the user.[10]

In four cases of technology competition the old technology had low extension opportunities. The Contact and Proximity tool architectures did not include a physical lens component (i.e., the "lens" that the energy passed through was air, not glass), so improving NA was not a possible extension mechanism. The Projection generation used a reflective lens that made it very difficult to increase NA beyond 0.167 (Henderson, 1995). The extension opportunity for these generations was also limited by the fact that they required the use of masks whose feature size was in a 1-to-1 ratio to the size of the features to be printed on the wafer. This created very high challenges for the mask production process and limited the potential for mask-based extension opportunities. By the time DUV 193-immersion was introduced in 2005 the extension opportunity for DUV 193 was severely limited by the physical limits imposed by the lens, mask and resist materials that had already been "stretched" for over two decades. In contrast, four generations had high extension opportunities. The G-line, I-line, DUV 248, and DUV 193 (at the time of the DUV 157 introduction) generations all benefitted enormously from extension opportunities due to improvements in lens design, lens manufacturing, in combination with new techniques for mask

[10] The technical term corresponding to the user's margin of error is the depth of focus (DOF). It is the maximum distance between two planes over which there are clear optical images. It can simply be considered as the range of focus errors the lithography process can tolerate and still provide acceptable results. Its relationship to energy wavelength and NA is given by $DOF = k_2 \times \lambda/(NA)^2$ where k_2 is a constant. An increase in NA, while improving resolution, lowers the DOF and makes the lithography process less robust. As compared to changes in wavelength, the squared NA term has a greater effect on the deterioration in the DOF and hence, limits the tool manufacturers' ability to improve the performance of the lithography technology by increasing the NA (cf. Henderson, 1995).

DOI: 10.1002/smj

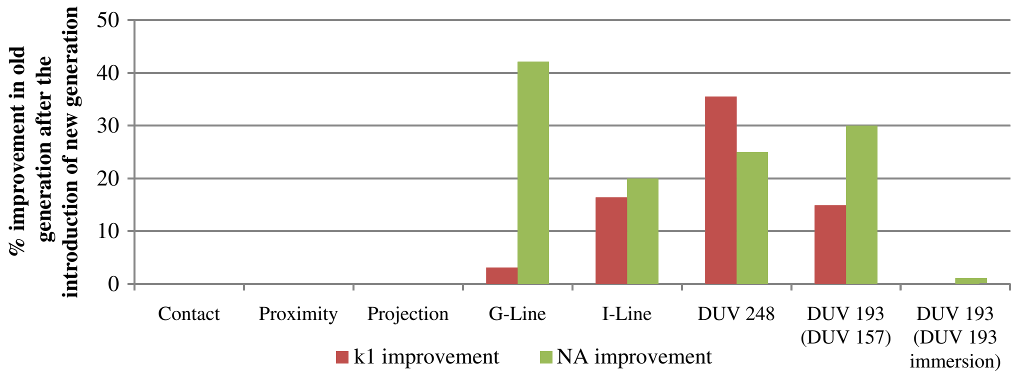


Figure 6. Magnitude of ecosystem extension opportunity for the different technology generations

design, mask manufacturing, and from the introduction and refinements of new resist chemistries.

To provide an objective measure of the extension opportunity for the different technology generations, we collected the published technical specifications for lithography tools sold by every major tool manufacturer between 1973 and 2009. Specifically, we collected data on resolution, numerical aperture (NA), wavelength (λ), and used these to derive the value of k1 for every tool model based on the Raleigh criterion. To quantify the extension opportunity in each generation that was due to improvements in the ecosystem, we computed the percentage increase in NA (extension due to lens component) and the percentage decrease in k1 (extension due to mask and resist complements) between the best old technology tool sold in the year that the new technology was introduced and the best old technology tool ever sold after that year. Figure 6 plots the percentage change in NA and k1 for the different technology generations.[11] Consistent with our understanding from the fieldwork, the characterization of old technology generations show an unambiguous distinction between generations with low extension opportunities (Contact, Proximity, Projection, DUV 193 following the introduction of DUV 193-immersion) and generations with high extension opportunities (G-line, I-line, DUV 248, DUV 193 following the introduction of DUV 157).[12]

Substitution dynamics: the joint consideration of emergence challenges and extension opportunities

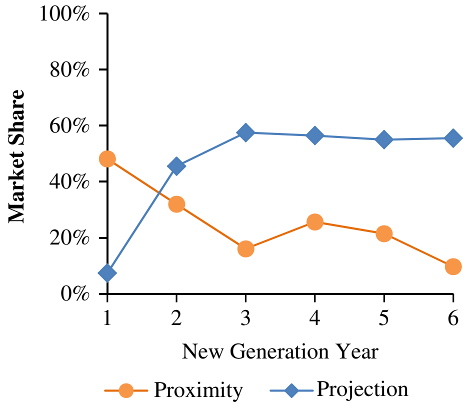
We have characterized each episode of technology competition according to the degree of ecosystem emergence challenge confronting the new technology and the ecosystem extension opportunity available to the old technology. We now consider these characterizations jointly to position each case of technology competition in its corresponding quadrant within our framework.

Four transitions were characterized by low ecosystem emergence challenge for the new technology and low ecosystem extension opportunity for the old technology (Quadrant 1 within our framework). Consistent with our arguments and as noted in the Appendix I, in all these cases the new technology generations (Proximity, Projection, G-line and DUV 193-immersion) achieved very rapid market adoption and were the fastest to substitute the old generation. Here, the average time to market dominance–the number of years from the year the new generation was commercialized to the year its annual market share surpassed that of the old generation–is three years.

[11] We note that improvements in k_1 were not entirely confined to innovations in mask and resist. Tool makers also contributed to the lowering of k_1 by changing the angle of light falling on the mask and hence, enhancing the contrast of the image. This innovation was known as Off-Axis Illumination (OAI). Although its introduction late in the G-Line generation was a non-negligible contributor to k1 improvement, our interviews suggest that its contribution was not as large as that from improvements in resist and mask. Because OAI was incorporated as a standard design element in every subsequent generation launch, it does not affect our measure of extension opportunity for any other generation.

[12] We note that the operationalization of extension opportunity is based on the *realized* extension of the old technology, not the *expected* opportunity for extension. In our interviews, industry experts were in unanimous agreement in characterizing the technology constraints that limited extension opportunities. After presenting our results, we explore the different modes through which this realization process has unfolded in the section titled "*The Role of Heterogeneity and Choice in Pacing Technology Substitution*."

DOI: 10.1002/smj



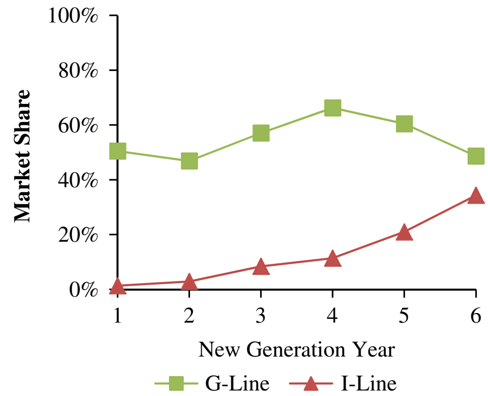


Figure 7. Difference in the pattern of substitution between low emergence challenge/low extension opportunity transition (proximity → projection) and low emergence challenge/high extension opportunity transition (G-line → I-line)

Three transitions were characterized by high ecosystem emergence challenge for the new technology and high ecosystem extension opportunity for the old technology (Quadrant 4). Consistent with our arguments, these new technology generations (DUV 248, DUV 193 and DUV 157) were the slowest to substitute the old generation.[13] A case in point for the high emergence challenge and high extension opportunity transition is the transition from I-line to DUV 248. It took DUV 248 11 years to achieve market dominance. The reduction in the wavelength from I-line's 365 to 248 nm resulted in technology bottlenecks requiring changes in the mask and resist complements. While developing a new mask material was difficult, the most significant emergence challenge in the ecosystem was with the resist. New resist chemistries had to be developed that were sensitive enough to absorb sufficient energy from the lower wavelength to cause the chemical reaction, yet specific enough so as not to react when energy impacted an adjoining molecule. It took several years before a new type of resist chemistry–chemically amplified resist–was developed to meet the commercialization requirements of the DUV 248 technology. In addition to the ecosystem emergence challenge faced by DUV 248 generation, the transition was further slowed by high extension opportunity for the I-line generation. This extension was achieved through increases in the NA of the lens component and through an innovation in the mask called Phase-Shift. Phase-Shift improved the resolution of the lithography technology by lowering k_1 and extending the I-line generation from its initial resolution limit of 500 nm to 350 nm (Terasawa *et al.,* 1989). This extension raised the performance threshold that the DUV 248 generation needed to cross in order to unlock its market potential and achieve dominance.

The transition from G-line to I-line was characterized by low ecosystem emergence challenge and high ecosystem extension opportunity (Quadrant 2). It took I-line seven years to dominate G-line. The intermediate pace of substitution for this transition is consistent with our arguments that the substitution dynamics within this quadrant would entail a robust competition between the old and the new technology until the old technology reaches its limits. Figure 7 illustrates this pattern of substitution and contrasts it with a typical case of fast substitution (Proximity substituted by Projection) observed for the low emergence challenge–low extension opportunity category. While Proximity's substitution by Projection is characterized by a rapid increase in Projection's share followed by market dominance and a rapid erosion of Proximity's share, G-line's substitution by I-Line is characterized by a much more gradual increase in I-line's share and a gradual decline in G-line's share. This case of the transition from G-line to I-line illustrates that low emergence challenge is not a sufficient condition for fast substitution. Specifically, the pace of substitution critically depends on the extension opportunity for the old technology.

Finally, the competition between Projection and E-beam in 1976, and then X-ray in 1978,

[13] In fact, the ecosystem emergence challenge for DUV 157 was so significant that, as of the time of this writing, it has yet to achieve any meaningful market penetration.

DOI: 10.1002/smj

was characterized by high ecosystem emergence challenge and low ecosystem extension opportunity (Quadrant 3). We expected that in this quadrant the old technology, despite having limited extension opportunity, would continue its market dominance until the emergence challenge with the new technology was resolved. However, once the emergence challenges are overcome, the new technology is expected to have a steep sales takeoff. Hence, in this regime, the old technology's dominance is fragile because its market position is based not in its own strength, but in the rival technology's weakness. The emergence challenges for E-beam and X-ray were both extremely high. In the absence of suitable resist and mask complements, neither generation could realize its technical performance advantage.[14] Consistent with our expectations, Projection dominated the market in the early years that followed the introduction of the new technologies. However, neither E-beam nor X-ray replaced Projection in later years. In this regard, the intermediate pace of substitution that we expected was not observed. This was due to the fact that the competition became a *multi* technology race. Specifically, the X-ray generation was commercialized two years after E-beam, and then the G-line generation (which had low emergence challenge) was commercialized before either E-beam's or the X-ray's emergence challenge had been overcome. As such, before they had had a chance to displace Projection, both technologies were themselves displaced. The technology competition for dominance rapidly shifted from the mode of old vs. new technology (Projection vs. E-beam) to old vs. *newer* technology (Projection vs. X-ray) to old vs. *newest* technology (Projection vs. G-line). One reason for the parallel efforts in these different technology generations is that they were pursued by different firms trying to exploit their own unique capabilities and positions.

In order to test whether the observed years to market dominance for the new technology within each of the quadrants follows the predicted ordering (years to dominance in Quadrant 1 < Quadrant 2 < Quadrant 4), we performed the Jonckheere-Terpstra non-parametric test for ordered alternatives. We were able to reject the null hypothesis of random ordering and our prediction was supported with the p-value of 0.009.

Regression analysis

To offer a more systematic analysis that also takes into account the substitution dynamics in Quadrant 3, we performed regression analysis. Our dependent variable is the new technology generation's annual share of total industry sales (in dollars) during the period of technology transition. The higher the new generation's annual market share during this period, the faster it substitutes the old generation. On average, it took a new generation six years to displace the old generation from its industry leadership. Hence, we perform our main analysis for the first six years of sales for each of the new technology generations. As a robustness check, we performed this analysis for the first four years and eight years as well.

In our analysis, we controlled for a number of generation- and industry-level effects that may influence the annual market share of the new technology generation. In order to account for the relative advantage of the new technology generation (Foster, 1986; Rogers, 2003), we control for the *price adjusted performance difference* (performance/price) between the new technology generation and the old technology generation. This measure was calculated based on average price and resolution capability of lithography tools sold in the given year. The primary form of technology competition existed between the new technology generation and the dominant old generation. However, some periods in the industry were characterized by the presence of multiple old technology generations that may have influenced the new generation's market share. We controlled for this effect with a count of the *number of generations* that captured at least 10 percent market share in the industry in a given year. We set 10 percent as the threshold to exclude cases of old technologies being used in small niche applications, such as Proximity generation tools being used in university labs 15 years after the generation has lost its dominance. The capabilities of the firm that pioneered the new technology generation, and the number of firms producing the new generation, may have also impacted market share. We accounted for these effects by categorizing whether the pioneering firm was an industry incumbent (*incumbent pioneer*), and by counting the number of firms that sold the new generation

[14] We note that while E-beam generation faced high ecosystem emergence challenge, it also suffered from lower throughput that made it less attractive for large scale mass production. We discuss this aspect of E-beam in detail in the Appendix I.

DOI: 10.1002/smj

Table 1. OLS regression estimates; dependent variable = generation share

	(1)	(2)	(3)	(4)	(5)	(6)
High emergence high extension		−0.328***	−0.262***	−0.316***	−0.453***	−0.283***
		(0.031)	(0.034)	(0.039)	(0.046)	(0.035)
Low emergence high extension		−0.235***	−0.174***	−0.159**	−0.386***	−0.161**
		(0.047)	(0.043)	(0.055)	(0.065)	(0.064)
High emergence low extension		−0.207***	−0.239***	−0.259***	−0.237***	
		(0.036)	(0.055)	(0.043)	(0.063)	
Price adjusted performance difference	2.338***	1.013*	1.868**	1.141**	0.340	2.139***
	(0.669)	(0.526)	(0.693)	(0.493)	(0.702)	(0.589)
Number of generations	−0.151***	−0.137***	−0.120***	−0.108***	−0.152***	−0.100**
	(0.031)	(0.024)	(0.033)	(0.025)	(0.038)	(0.038)
Incumbent pioneer	−0.086	−0.006	0.027	−0.049 *	0.040	−0.017
	(0.090)	(0.018)	(0.020)	(0.024)	(0.032)	(0.019)
Number of firms	0.019	0.007 *	0.008	0.009	0.014 *	0.001
	(0.016)	(0.004)	(0.005)	(0.006)	(0.007)	(0.005)
Old generation tenure	−0.002	−0.010***	−0.009**	−0.006	−0.016**	−0.009**
	(0.021)	(0.003)	(0.003)	(0.004)	(0.006)	(0.003)
Newer generation availability	−0.037	−0.145**	−0.196***	−0.161***	−0.014	−0.211***
	(0.112)	(0.062)	(0.055)	(0.049)	(0.085)	(0.046)
Year trend	0.019	0.039**	0.052*	0.023	0.050**	0.053***
	(0.023)	(0.014)	(0.026)	(0.014)	(0.021)	(0.015)
Constant	0.316*	0.554***	0.375***	0.497***	0.736***	0.356**
	(0.148)	(0.069)	(0.092)	(0.078)	(0.103)	(0.119)
Observations	59	59	40	75	59	47
R-squared	0.48	0.82	0.79	0.75	0.81	0.84

Omitted category is low emergence & low extension. Robust standard errors in parentheses, clustered by technology generation.
*Significant at 10 percent; **significant at 5 percent; ***significant at 1 percent.

in a given year (*number of firms*). The pace of substitution may also be impacted by the differences in the semiconductor manufacturers' switching costs across the different technology transitions. The longer the generation had been used by semiconductor manufacturers, the greater may be their switching cost. We controlled for this possibility through the variable, *old generation tenure*, which is the number of years the old generation has been sold in the industry. The new generation's market share may also be impacted if the period of observation is characterized by the availability of a newer technology generation with low emergence challenge. This effect is controlled for by the dummy variable*, newer generation availability*. Finally, we control for the *yearly trend* in the market share of a given generation during the period of substitution.

Table 1 reports the estimates from the OLS regression analysis. The standard errors are clustered by technology generation to account for the non-independence of observations within a generation. Model 1 is the baseline model with control variables, and Model 2 includes the effects of the different substitution regimes with the omitted category of low emergence challenge and low extension opportunity. The results from the baseline model are consistent with our expectations. The greater the price adjusted performance difference between the new and the old generation, the greater is the new generation's market share. The market share of the new generation decreases with the number of generations in the industry. The coefficients for number of firms, old generation tenure and newer generation availability exhibit expected signs but are significant only in the full model.

The coefficients for all three substitution regimes (high emergence and high extension; low emergence and high extension; high emergence and low extension) are negative and significant supporting our prediction that the technology transitions characterized by low emergence challenge and low extension opportunity (omitted category) exhibit the fastest substitution. The magnitudes of the three coefficients are in the expected order with the coefficient for the high emergence and high extension category having the highest magnitude. A statistical comparison of the coefficients for the three regimes reveals that the coefficient for the high emergence-high extension category is significantly different from the low emergence-high extension category (p-value of 0.002) and the high emergence-low extension category (p-value of

0.057). Hence, as hypothesized, the technology transitions characterized by high emergence challenge and high extension opportunity have the slowest pace of substitution. The coefficients for the low emergence challenge and high extension opportunity category, and that for the high emergence challenge and low extension opportunity category supported our prediction of intermediate pace of substitution with the difference between the two coefficients being statistically insignificant (p-value of 0.71).[15]

We also performed a number of additional robustness checks in order to ensure that the regression estimates are not sensitive to our variable specifications and that our inferences are not subjected to alternative explanations. In Models 3 and 4, we report the estimates using data from the first four and eight years of sales for each of the new technology generations respectively. We also checked that our results are not sensitive to the presence of multiple technology generations. In Model 5 we used an alternative measure of generation share by explicitly considering head-to-head technology competition between the new and the dominant old generation. Here generation share is based on only the sales of the new generation and the old dominant generation. Finally, since X-ray and E-beam are non-optical generations and may be characterized as having higher levels of complexity and incompatibility with semiconductor manufacturers' past experiences (Rogers, 2003), we excluded them from the estimates reported in Model 6. The results were robust to these analyses. The only exception being that the coefficient for the high emergence and high extension category was not significantly different from that of the high emergence and low extension category in Models 3 and 4. This is probably because the two new generations in the high emergence and low extension category (X-ray and E-beam) never replaced the old technology. File S1, that is available as an online supplement, examines additional alternative explanations for our results with regards to users' capabilities, preferences and heterogeneity and also examines an alternative characterization of the emergence challenge measure.

[15] An important issue pertaining to the regression analysis is that the statistical power of the test, owing to small sample size, may be too low to yield any meaningful insights. We performed a power analysis using the "powerreg" procedure in STATA and found that because of our large effect size, our analysis has very high power (approaching 1.00; almost a 100 percent probability of rejecting the null hypothesis if it is false).

THE ROLE OF HETEROGENEITY AND CHOICE IN PACING TECHNOLOGY SUBSTITUTION

Our theory and analysis are at the level of technology. Technological progress, however, does not happen on its own, but rather is shaped by the choices and efforts of the heterogeneous participants that compose the ecosystem. In studying the patterns of technological change in the semiconductor lithography equipment industry we were able to uncover three different modes by which industry participants influenced the pace of substitution.

First, in some cases, the pace of substitution was slowed by the "last gasp" effort by some incumbent firms to prolong the life of their old technology generation. We observed such a pattern in the case of the G-line generation after the emergence of the I-line generation. Nikon was the first firm to commercialize the I-line generation in 1985. In contrast, Canon, another incumbent, chose to focus its R&D efforts on extending the G-line generation through an increase in the size of the lens. Canon introduced its I-line tools years after Nikon, only when further extending the performance of the G-line generation was deemed unfeasible. This kind of last gasp effort to reinvigorate performance improvement in mature technologies has been observed in other technologies such as sailing ships (Harley, 1971), pond ice harvesting (Utterback, 1994), typesetters (Tripsas, 2008). The main driver here is the differences in innovation incentives and strategies among the different focal firms.

Second, there were instances in which R&D efforts initiated to resolve the emergence challenges in the new technology generation triggered the discovery of solutions that served to also extend the old technology generation. For example, one of the key challenges confronting the DUV 157 generation was the development of new lens materials with high concentration of Calcium Fluoride that would allow for transmission of low wavelength light. The R&D efforts to resolve this challenge also helped firms to make better lenses for the old DUV 193 generation, extending its performance. Similarly, the resist development for newer generations consistently helped in improving the chemical composition of the resist used with the older generations. Furr and Snow (2014) document such a "spillback" effect in the automotive industry where the development of the new electronic fuel injection technology helped to extend the performance of

DOI: 10.1002/smj

the old carburetor technology. They also found that the extent of spillback realized by individual firms varied according to their specific level of investment in the new technology. Taylor (2010) explores the intra-organizational mechanisms through which such spillbacks take place.

Finally, we observed a third mechanism that has not been previously characterized in the literature. We found instances where participants across the lithography ecosystem (competing firms, users, suppliers and complementors) engaged in a collective "last resort" effort to extend the old technology due to a collective inability to overcome the emergence challenges in the new technology. For example, when it was evident that the emergence of DUV 193 in 1996 would be significantly delayed due to the challenges in the resist, which would mean that the generation would not be able to achieve its planned introduction at 180 nm resolution, lithography equipment firms, semiconductor manufacturers, materials and other equipment manufacturers collectively revised their development emphasis to extend the performance of the incumbent DUV 248 generation through a combination of improvements in the tool, lens, resist and mask.[16]

Each of the three modes represents a distinct type of firm-level actions in shaping the pace of substitution. In the case of last gasps, some firms aggressively pursue the new technology while other firms maximize the performance from the old technology. In the case of spillbacks, R&D efforts toward the new technology help firms to also extend the old technology. In the case of last resort, firms across the ecosystem collectively reallocate their resources to extend the old technology when met with significant bottlenecks in the new technology. These different modes of firms' actions illustrate how technology transitions are shaped by a rich set of collaborative and competitive interactions among heterogeneous firms across the ecosystem that extend beyond technology constraints and opportunities.

DISCUSSION AND CONCLUSION

This study sheds light on the forces that determine the pace of technology substitution by presenting a framework that identifies when and how technology competition is resolved. It considers the relative balance between the emergence challenges confronting the new technology and the extension opportunities available to the old technology as a key governor of these dynamics. Underlying this approach is a view that, in order to understand progress in a given technology, we need to take into account progress in the surrounding ecosystem in which the focal technology is embedded (e.g., Hughes, 1983; Rosenberg, 1976). We apply this framework to study 10 episodes of technology competition that have occurred in the semiconductor lithography equipment industry from 1972 to 2009. We find that the observed dynamics are fully consistent with our predictions. The framework also helps to clarify the puzzling variance in transition paths observed in the industry despite the existence of structural similarities across the different cases of technology competition.

In addition to these core findings, our in-depth study of the semiconductor lithography equipment industry sheds light on the different modes of firm-level actions that shaped the observed pattern of substitution. We found that at times, the pace of substitution was slowed due to "last gasp" efforts by some firms to maximize the value that they could capture from the old technology while other firms aggressively pursued the new technology. In other instances, R&D efforts in the new technology created a "spillback" effect and helped firms to also improve the performance of the old technology. Finally, we found evidence of heterogeneous actors across the ecosystem engaging in a collective "last resort" effort to extend the old technology when met with significant emergence challenges in the new technology.

[16] Since the 1990s, progress in semiconductor manufacturing has been coordinated according to collectively defined technology roadmaps, most recently under the auspices of International Technology Roadmap of Semiconductors (ITRS). The roadmap is regarded as a very good indicator of the R&D resource allocation across the different technology generations. The 1991 roadmap, in expectation that the DUV 193 generation would launch in 1996 with a resolution of 250 nm, specified that the performance of DUV 248 would peak at 250 nm and that this limit would be reached by 1996 (Micro Tech 2000, 1991). As emergence challenges delayed the launch of the DUV 193 generation, the 1997 and 1999 roadmaps revised the expected performance limit of DUV 248 to 150 nm by 1999 and 130 nm by 2001 respectively. These revisions signaled a collective agreement by actors across the lithography ecosystem to reallocate substantial development resources to extend the performance of DUV 248 (and so away from the DUV 193 effort) as a last resort for sustaining the trajectory guided by Moore's law. As a result of these investments the DUV 248 generation was extended to an ultimate resolution performance of 130 nm (in 2001), which sustained the industry until the emergence challenges of DUV 193 were resolved in 2002.

DOI: 10.1002/smj

Approaching substitution as an outcome of the interplay between extension of the old technology and emergence of the new helps clarify observations that highly discontinuous changes at one level of a technology hierarchy can be invisible at other levels (e.g., Funk, 2008). In the context of semiconductor industry, we observe that despite many misses and delays in the commercialization of new lithography technologies, semiconductor manufacturers continued to make steady progress along Moore's law. Whenever semiconductor manufacturers were confronted with the prospect that their lithography tool suppliers would be unable to meet the industry roadmap requirements, they proactively guided resources to find ways of bridging the gap with the old technology until a new technology solution became available. This "last resort" reallocation of efforts had significant impact on the returns to lithography equipment firms' investments in both the old and new technology generations.

By focusing on ecosystem dynamics that underlie technology transitions, we are able to draw a link between two of the most fundamental life-cycle patterns identified in the innovation community: the S-curve profile of technology improvement and the S-curve profile of technology adoption. While the primary use of the technology S-curve has been to understand a given technology's performance take-off and limit so as to guide resource allocation into existing and new technologies, the primary use of the adoption S-curve has been to understand the given technology's market potential and sales take-off so as to guide market expectations and commercialization strategies for new technologies (e.g., Agarwal and Bayus, 2002; Bass, 2004; Fisher and Pry, 1972). Studies of adoption patterns focus on the dynamics of the demand environment (e.g., Griliches, 1957; Rogers, 2003), but most often assume that both the old and new technologies are static (Hall, 2004). Conversely, studies of technology evolution focus almost exclusively on the relative performance position of the new and old technologies, and tend to assume that a technology with superior performance will automatically take over a given market. In doing so, they have tended to overlook the role of demand-side adoption in determining the pace and extent of technology substitution (Adner, 2004).

Our paper informs and clarifies the relationship between these two fundamental life-cycle patterns. Specifically, we note that adoption and substitution are intimately linked–a user's decision to adopt the new technology is, implicitly, a decision not to adopt the incumbent technology. Hence, what looks like adoption from the new technology's perspective looks like substitution from the old technology's perspective. They are two sides of the same coin. For this reason, understanding the adoption of a new technology requires not only a detailed demand side analysis of consumer heterogeneity and information diffusion mechanism (e.g., Rogers, 2003) but also consideration of the supply side factors that govern the technological progress within an ecosystem. As with substitution, the pace of adoption will be a function of the dynamics of emergence in the new technology's ecosystem and the dynamics of extension in the incumbent technology's ecosystem. The emergence of a new technology may require major improvements in the complements for the average user to derive the full benefits, and this may slow the adoption rate. Similarly, the extension of a new technology through improvements in components, architectures or complements may also slow the adoption of the new technology. The framework that we develop in this study is able to explain why new superior technologies may face slow market adoption and why seemingly geriatric technologies may continue to dominate.

In this regard, our approach has some predictive implications. Historically, S-curve analysis has focused on the performance profile of a stand-alone technology. One problem with the S-curve is that the time to performance take-off is notoriously difficult to specify ex-ante. Focusing on the broader technological ecosystem expands the analysis beyond the time to take-off for the single focal technology, to include the performance profile of the entire ecosystem on which the focal technology's performance as realized by the user depends on. This expansion of the consideration set can improve our ability to make predictions. In an interdependent system, the realized performance by the user cannot take off until every element is ready. For this reason, the most challenged element acts as a bottleneck for the whole system. This has an important implication: the take-off of the focal technology's technical performance will be realized by the user only if the focal technology is itself the bottleneck of the system. If the focal technology is not the bottleneck, overcoming its own development challenge will not resolve the ecosystem's emergence challenge.

DOI: 10.1002/smj

Another critique of the S-curve representation of technology evolution is the inconsistency between the theorized smooth S-shaped and the observed irregular pattern of technology evolution (Sood and Tellis, 2005). By taking an ecosystem view of technology and by considering technology transitions as an interaction between new technology's ecosystem emergence challenge and old technology's ecosystem extension opportunity, our approach helps to provide some micro-foundations for the observed differences in the pattern of technology evolution.

Although this study is conducted at the level of technologies, its findings have a number of implications for firms. First, we propose that rather than viewing technology substitution as the race between focal technologies, firms view them as the race between technology ecosystems. This view can inform the way in which firms choose to allocate their valuable R&D resources according to the likely benefits from investing early in the new technology or continuing to focus on the old technology (e.g., Adner and Kapoor, 2010; Furr and Snow, 2014). Second, the explicit consideration of technological interdependencies and asymmetries in the rate of advance among the different ecosystem elements can help firms consider various modes of coordination ranging from vertical integration to collaborative alliances (e.g., Hughes, 1983; Kapoor, 2013).

The logic of our framework can also help firms in their resource allocation choices between technology development initiatives, investment in ecosystem complements, and investment in complementary assets such as those in manufacturing, marketing, sales and distribution. In Quadrant 1, with the old technology stagnant and the new technology unhampered, firms should aggressively invest in the new technology, following the traditional prescriptions for embracing change to ward off the winds of creative destruction. In contrast, when emergence challenges for the new technology are low and the extension opportunities for the old technology are high (Quadrant 2), firms with established positions in the old technologies can afford to maintain their investment focus on the old technology for a longer time, knowing that the technologies will coexist for an extended period. Alternatively, when emergence challenges for the new technology are high and the opportunity for extending the old technology is low (Quadrant 3), firms may be better off allocating a greater proportion of resources towards the development of complementary assets than those in the development of either the old or new technologies. Most favored here would be specialized complementary assets that translate well across both technologies (e.g., Mitchell, 1989; Rothaermel and Hill, 2005). Finally, in Quadrant 4, firms may be best served in investing in ecosystem complements that enhance the performance of the old technology or offer an advantage in unleashing the potential of the new technology. Here, our paper extends the literature exploring the role of firms' complementary assets during periods of technological change (e.g., Mitchell, 1989; Teece, 1986). For example, Rothaermel and Hill (2005) find that industry- and firm- level heterogeneity in the face of competence destroying technologies are impacted by the nature of the complementary assets (generic versus specialized) required to commercialize the new technology. We complement their work by focusing on complementary technologies in the ecosystem that surround both the old and the new focal technologies, and by exploring transitions in a setting where technological discontinuities are not competence destroying.

Finally, our study also provides some useful insights for policy. While we suggest that the rate of technology substitution depends on the relative difference in the emergence challenges and extension opportunities that confront the new and old technologies, we note that three of the four types of substitution regimes are characterized by significant technical progress fueled either by the old or the new technology. However, when emergence challenges are high and extension opportunities are low, industries may face stagnation. In Quadrant 3, policy makers may have to create additional incentives and mechanisms through which the emergence challenges with the new technology can be overcome in a timely manner. In addition, a robust technology policy may also consider alternative new technologies that may be more likely to overcome emergence challenges.

Although our focus is on substitution dynamics at the level of technologies, it is clear that these dynamics result from the aggregation of decisions and actions of individual firms. We have taken the magnitude of the technology challenges and opportunities as relatively exogenous but note that the resolution of these challenges and realization of these opportunities are highly endogenous, depending on the resources, capabilities, motivation and strategies of firms across the ecosystem. Our elucidation of how ecosystem dynamics at the

DOI: 10.1002/smj

technology-level influence the pace of substitution gives rise to a new set of research questions at the firm-level. For example, while scholars have studied the relative advantages and disadvantages of incumbents during technology transitions according to differences with respect to competences (Tushman and Anderson, 1986), complementary assets (Tripsas, 1997), cognition (Henderson and Clark, 1990) and customer dependencies (Christensen, 1997), our study raises an additional dimension of difference that is rooted in the interaction between the old and new technology ecosystems. We expect that the performance difference between incumbents and entrants would vary across the different quadrants within our framework. Similarly, while numerous studies have explored how firms' resources and capabilities influence their entry and performance outcomes in new technologies (e.g., Franco *et al.,* 2009; King and Tucci, 2002; Mitchell, 1989, 1991), our study suggests that a more holistic evaluation would consider not only how firms compete in new technologies but also how they transition out of the old and into the new with an explicit recognition of the interdependencies in the ecosystem.

As in any investigation of an open system, a critical choice regards where to draw the boundaries of the analysis (Scott, 2003). In this study we limit our attention to the technological system (e.g., Constant, 1980; Hughes, 1983; Rosenberg, 1976). Within this system, we focus on the external elements that link directly to the focal technology–the components and complements that combine with the focal technology to determine realized performance. Outside the scope of our study, but a promising direction for future work, is the host of non-technological social and economic factors (e.g., coordination costs; distribution of power among participants; industrial organization across sectors; identity; institutional, regulatory, and societal concerns) that may play a critical role in determining the timing and outcomes of technology transitions. We expect exploration of these dynamics to be particularly fruitful and hope our study has paved the way for scholars to push this research agenda forward.

ACKNOWLEDGMENTS

We thank Rajshree Agarwal, Phil Anderson, Mary Benner, Constance Helfat, Peter Golder, Dan Levinthal, Paul Rosenbaum, Lori Rosenkopf, Dan Snow, Mary Tripsas, Christophe Van den Bulte, Peter Zemsky, Arvids Ziedonis, and seminar participants at Columbia, HBS, INSEAD, MIT, USC, UNC, Wharton, and the attendees at the 2010 West Coast Research Symposium for their helpful comments. We are grateful to editor Rich Bettis and the many anonymous reviewers for their guidance and feedback. We especially thank VLSI Research for their generosity in sharing their data with us and the many industry professionals who provided us with valuable insights.

REFERENCES

Abernathy WJ, Clark KB. 1985. Innovation: mapping the winds of creative destruction. *Research Policy* **14**: 3–22.

Adner R. 2002. When are technologies disruptive: a demand-based view of the emergence of competition. *Strategic Management Journal* **23**: 667–688.

Adner R. 2004. A demand-based perspective on technology life cycles. In *Business Strategy Over the Industry Lifecycle: Advances in Strategic Management* (Volume 21), Baum J, McGahan A (eds). Elsevier: Oxford, UK; 25–43.

Adner R. 2006. Match your innovation strategy to your innovation ecosystem. *Harvard Business Review* **84**(4): 98–107.

Adner R. 2012. *The Wide Lens: A New Strategy for Innovation*. Penguin: London, UK.

Adner R, Kapoor R. 2010. Value creation in innovation ecosystems: how the structure of technological interdependence affects firm performance in new technology generations. *Strategic Management Journal* **30**(3): 306–333.

Adner R, Levinthal DA. 2001. The emergence of emerging technologies. *California Management Review* **45**(1): 50–66.

Agarwal R, Bayus B. 2002. The market evolution and sales takeoff of product innovations. *Management Science* **48**(8): 1024–1041.

Anderson PC, Tushman ML. 1990. Technological discontinuities and dominant designs. *Administrative Science Quarterly* **35**(4): 604–633.

Ansari S, Garud R. 2009. Inter-generational transitions in socio-technical systems: the case of mobile communications. *Research Policy* **38**(2): 382–392.

Bass FM. 2004. Comments on "A new product growth for model consumer durables the bass model". *Management Science* **50**(12S): 1833–1840.

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

Cusumano MA, Mylonadis Y, Rosenbloom RS. (1992). Strategic maneuvering and mass-market dynamics: The triumph of VHS over Beta. *Business history review*, **66**(01), 51–94.

DOI: 10.1002/smj

Christensen CM, Rosenbloom RS. 1995. Explaining the attacker's advantage: technological paradigms, organizational dynamics, and the value network. *Research Policy* **24**(2): 233–257.

Constant EW. 1980. *The Origins of the Turbojet Revolution*. Johns Hopkins University Press: Baltimore, MD.

Dranove D, Gandal N. 2003. The Dvd-vs.-Divx standard war: empirical evidence of network effects and preannouncement effects. *Journal of Economics & Management Strategy* **12**(3): 363–386.

Ethiraj SK. 2007. Allocation of inventive effort in complex product systems. *Strategic Management Journal* **28**(6): 563–584.

Farrel J, Saloner G. 1986. Installed base and compatability: innovation, product preannouncements, and predation. *American Economic Review* **76**(5): 940–955.

Fisher J, Pry R. 1972. A simple substitution model of technological change. *Technological Forecasting and Social Change* **3**: 75–88.

Foster R. 1986. *Innovation: The Attacker's Advantage*. Summit Books: New York.

Franco AM, Sarkar MB, Agarwal R, Echambadi R. 2009. Swift and smart: the moderating effects of technological capabilities on the market pioneering-firm survival relationship. *Management Science* **55**(11): 1842–1860.

Funk J. 2008. Components, systems, and technological discontinuities: lessons from the IT sector. *Long Range Planning* **41**(5): 555–573.

Furr NR, Snow D. 2014. Inter-generational hybrids: spillbacks, spillforwards, and adapting to technology discontinuities. *Organization Science*. http://pubsonline.informs.org/doi/abs/10.1287/orsc.2014.0930 (26 February 2015).

Griliches Z. 1957. Hybrid corn: an exploration in the economics of technological change.

Hall B. 2004. Innovation and diffusion. In *The Oxford Handbook of Innovation*, Fagerberg J, Mowery D, Nelson RR (eds). Oxford University Press: Oxford, UK; 459–485.

Harley CK. 1971. The shift from sailing ships to steamships, 1850–1890: a study in technological change and its diffusion. In *Essays on a Mature Economy: Britain after 1840* (Volume xv), McCloskey DN (ed). Princeton University Press: Princeton, NJ; 439 pp.

Helfat CE, Klepper S. 2007. Firm and industry evolution and entrepreneurship (FIVE project): data overview (November 2, 2007). In *2007 Kauffman Symposium on Entrepreneurship and Innovation Data*. Available at SSRN: http://ssrn.com/abstract=1022116. (accessed 10 January 2008).

Henderson R. 1993. Underinvestment and incompetence as responses to radical innovation: evidence from the photolithographic alignment equipment industry. *RAND Journal of Economics* **24**(2): 248–270.

Henderson RM. 1995. Of life cycles real and imaginary: the unexpectedly long old age of optical lithography. *Research Policy* **24**(4): 631–643.

Henderson RM, Clark KB. 1990. Architectural innovation: the reconfiguration of existing product technologies and the failure of established firms. *Administrative Science Quarterly* **35**(1): 9–30.

Hibbs M, Kunz R, Rothschild M. 1995. 193-nm lithography at MIT Lincoln lab. *Solid State Technology* **38**: 69–78.

Hughes TP. 1983. *Networks of Power: Electrification in Western Society 1880–1930*. Johns Hopkins University Press: Baltimore, MD.

Iansiti M. 1998. *Technology Integration: Making Critical Choices in a Dynamic World*. Harvard Business School Press: Boston, MA.

Iansiti M, Levien R. 2004. *The Keystone Advantage: What the New Dynamics of Business Ecosystems Mean for Strategy, Innovation, and Sustainability*. Harvard Business School Press: Boston, MA.

Iyengar R, Van den Bulte C, Valente T. 2011. Opinion leadership and social contagion in new product diffusion. *Marketing Science* **30**(2): 195–212.

Kapoor R. 2013. Collaborating with complementors: what do firms do? *Advances in Strategic Management* **30**: 3–25.

Kapoor R, McGrath PJ. 2014. Unmasking the interplay between technology evolution and R&D collaboration: evidence from the global semiconductor manufacturing industry, 1990–2010. *Research Policy* **43**(3): 555–569.

Katz ML, Shapiro C. 1985. Network externalities, competition, and compatibility. *American Economic Review* **75**(3): 424–440.

King AA, Tucci CL. 2002. Incumbent entry into new market niches: the role of experience and managerial choice in the creation of dynamic capabilities. *Management Science* **48**(2): 171–186.

Levinthal DA. 1998. The slow pace of rapid technological change: gradualism and punctuation in technological change. *Industrial and Corporate Change* **7**(2): 217–247.

McGahan AM. 2004. *How Industries Evolve: Principles for Achieving and Sustaining Superior Performance*. Harvard Business School Press: Boston, MA.

Mitchell W. 1989. Whether and when? Probability and timing of incumbents' entry into emerging industrial subfields. *Administrative Science Quarterly* **34**(2): 208–230.

Mitchell W. 1991. Dual clocks: entry order influences on incumbent and newcomer market share and survival when specialized assets retain their value. *Strategic Management Journal* **12**(2): 85–100.

Moore JF. 1993. Predators and prey: a new ecology of competition. *Harvard Business Review* **71**(3): 75–86.

National Advisory Committee on Semiconductors. 1991. *MICROTECH 2000 Workshop Report*. U.S. Government Printing Office: Washington, DC.

Rogers EM. 2003. *Diffusion of Innovations*. The Free Press: New York.

Rosenberg N. 1976. *Perspectives on Technology*. Cambridge University Press: Cambridge, UK.

Rosenberg N. 1982. *Inside the Black Box: Technology & Economics*. Cambridge University Press: Cambridge, UK.

Rothaermel FT, Hill CW. 2005. Technological discontinuities and complementary assets: a longitudinal study of industry and firm performance. *Organization Science* **16**(1): 52–70.

Scott WR. 2003. *Organizations: Rational, Natural and Open Systems*. Upper Saddle River, NJ: Prentice-Hall. 5th ed.

Snow D. 2008. Extraordinary efficiency growth in response to new technology entry: the carburetor's "Last Gasp". Harvard Business School Working Paper #1668643.

Sood A, Tellis GJ. 2005. Technological evolution & radical innovation. *Journal of Marketing* **69**(3): 152–168.

Stremersch S, Tellis GJ, Franses PH, Binken JLG. 2007. Indirect network effects in new product growth. *Journal of Marketing* **71**(3): 52–74.

Taylor A. 2010. The next generation: technology adoption and integration through internal competition in new product development. *Organization Science* **21**: 23–41.

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

Terasawa T, Hasegawa N, Kurosaki T, Tanaka T. 1989. 0.3-micron optical lithography using a phase-shifting mask. *Proceedings of the International Society for Optical Engineering* **1088**: 25–33.

Tripsas M. 1997. Unraveling the process of creative destruction: complementary assets and incumbent survival in the typesetter industry. *Strategic Management Journal* **18**(6): 119–142.

Tripsas M. 2008. Customer preference discontinuities: a trigger for radical technological change. *Managerial and Decision Economics* **29**(2–3): 79–97.

Tushman ML, Anderson P. 1986. Technological discontinuities and organizational environments. *Administrative Science Quarterly* **31**(3): 439–465.

Utterback JM. 1994. *Mastering the Dynamics of Innovation: How Companies can Seize Opportunities in the Face of Technological Change*. Harvard Business School Press: Boston, MA.

APPENDIX I: SEMICONDUCTOR LITHOGRAPHY EQUIPMENT GENERATIONS AND THEIR EMERGENCE CHALLENGES

1. Contact Printers (First Generation)

Contact printing was the earliest and the simplest of the lithography technologies to be commercialized in 1962. The modifier "contact" corresponds to the fact that the mask and the semiconductor wafer were in direct contact with each other. The contact printers in the 1960s included a stage system for securing the mask and the wafer, and an alignment unit that ensured that the patterns from the mask were accurately transferred to the wafer. They used a mercury lamp for their light source. Contact printers did not incorporate a lens.

2. Proximity Printers (Year of Introduction: 1972; Years to Market Dominance: 2)[17]

A primary disadvantage of contact printing was low process yield. This was due to the damage inflicted on the mask and wafers as they were repeatedly brought in and out of physical contact with each other during the lithography process. Proximity printing offered a way to overcome this problem. With proximity printing the mask and the wafer were separated by a tiny gap. As discussed in detail by Henderson and Clark (1990), this generational transition can be characterized as an architectural innovation and imposed significant challenges on the focal lithography equipment firms. The transition to proximity printing did not, however, present major challenges to any of the complements in the lithography ecosystem.

3. Projection Scanners (Year of Introduction: 1973; Years to Market Dominance: 2)

The continuous need of semiconductor manufacturers to reduce circuit dimensions and get better manufacturing yields led to the introduction of another architectural innovation. Projection scanners introduced the use of lens systems as a component in lithography tools. The lens system was composed of a series of reflective mirrors which allowed the image of the mask to be transferred to the wafer. While the development of projection scanners entailed the development of the lens system, the commercialization of this generation did not pose any significant ecosystem emergence challenge. Although masks had to be etched with correspondingly smaller geometries and greater accuracy which required that mask makers switch from using step and repeat cameras to using the more accurate electron beam systems in their production process, this change was somewhat incremental.

4. E-beam Writers (Year of Introduction: 1976; Years to Market Dominance: –)

Electron-beam (E-beam) technology involves patterning the resist on the semiconductor wafer directly using electron beams that follow a pre-programmed pattern, thereby eliminating the need for a mask. Since a pre-programmed electron beam travels across the wafer to achieve very low

[17] Market dominance is defined as the first year in which the new generation exceeds the annual market share of the old (previously dominant) generation. In generations marked (–) the new technology did not achieve dominance.

DOI: 10.1002/smj

resolution, the time required to complete a single wafer can be as high as 10 hours, which is almost an order of magnitude longer than the alternate optical lithography technologies. The major ecosystem emergence challenge posed by E-beam technology was the development of new resist chemistries that work with the emitted electrons and that would allow for sufficiently small geometries; that is, a resist in which exposure to electron beam would trigger a chemical reaction only in the molecules that were directly exposed to the energy, without setting off a reaction in adjacent molecules.

5. X-ray Printers (Year of Introduction: 1978; Years to Market Dominance: –)

The use of X-rays for lithography was proposed due to their very low wavelength of less than 10 nm. In the early 1970s, X-ray lithography was developed as a simple proximity imaging system. As in proximity printing, the radiation from an X-ray source is transmitted through the mask onto the resist. X-ray lithography's very low wavelength created substantial ecosystem emergence challenges. The most significant challenge included the development of a suitable mask that would allow X-ray to accurately pass through and achieve the desired chemical reaction with the resist.

6. G-line (Year of Introduction: 1978; Years to Market Dominance : 5)

The G-line stepper introduced two key modifications. First, light was projected through the mask on to the wafer using a refractive lens system (as opposed to the reflective lens system used in projection scanners). Second, the light was projected on only a part of the wafer at any one time; the mask was shifted across the wafer in steps, such that multiple exposures are made across the wafer to complete the lithography process. This significantly eased the challenge of mask making as the circuit patterns on the mask could now be 10× or 5× of the dimensions that need to be printed on to the wafer. While resist suppliers had to develop new novolac-based materials in order to achieve the miniaturization objectives of the G-line generation, it did not present a major ecosystem emergence challenge.

7. I-line (Year of Introduction: 1985; Years to Market Dominance: 7)

I-line steppers, introduced in 1985, used light with a wavelength of 365 nm to improve over the resolution achievable with the G-line generation. The generation required development of a lens system that would transmit light at the lower wavelength. However, the ecosystem complements of mask and resist required only incremental changes for the transition from the G-line to the I-line generation.

8. DUV 248 (Year of Introduction: 1988; Years to Market Dominance: 11)

The DUV 248 generation entailed a further reduction in wavelength into the deep ultraviolet (DUV) spectrum at 248 nm. The reduction created a high degree of ecosystem emergence challenge requiring fundamental changes in the mask and the resist. The challenges imposed on mask makers were overcome by changing the mask material from soda lime glass to quartz in order to provide improved transmission of the 248 nm wavelength. This, in turn, required major changes to the mask manufacturing process. The existing novolac resists could not absorb enough energy from the new wavelength to cause an adequate chemical reaction. As a result, new chemically-amplified resists had to be developed for semiconductor manufacturers to create fine circuits using the new lithography technology.

9. DUV 193 (Year of Introduction: 1996; Years to Market Dominance: 11)

The industry's drive towards finer resolutions continued when tools using the 193 nm wavelength were introduced in 1996. As was the case for the DUV 248 nm technology, the very low light wavelength required major developments in the resist formulation so that this new generation could create value for users. With the change to the 193 nm wavelength, the existing resists, which were engineered to react to the 248 nm wavelength, were no longer adequate to the task–a new generation of chemically amplified resist needed to be developed.

10. DUV 157 (Year of Introduction: 1998; Years to Market Dominance: –)

The subsequent attempts to achieve greater miniaturization resulted in the development of tools using the 157 nm wavelength. The reduction in wavelength created transmittance problems and required that the mask substrate and the

DOI: 10.1002/smj

pellicle materials to be completely redeveloped to effectively transmit the low wavelength light.

11. DUV 193-immersion (Year of Introduction: 2005; Years to Market Dominance: 4)

The transition to DUV 193-immersion was enabled by an architectural innovation in which liquid (instead of air) is used as a medium between the lens and the wafer. The generation required major efforts by tool producers to manage the flow of liquid in this complex architecture. The transition did not, however, present major innovation challenges to any other elements in the lithography ecosystem. As the wavelength of light did not change, the new generation was able to reuse the existing resist and mask with relatively minor improvements.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article:

File S1. Innovation ecosystems and the pace of substitution: re-examining technology S-curves.

DOI: 10.1002/smj