Scaling Moore's Wall: Existing Institutions and the End of a Technology Paradigm

Submitted in partial fulfillment of the requirements for

the degree of

Doctor of Philosophy

in

Engineering and Public Policy

Hassan N. Khan

B.S., Chemical Engineering, University of California, Berkeley

Carnegie Mellon University
Pittsburgh, PA

December, 2017

Acknowledgements

The journey to completing and submitting a dissertation has been far different than what I originally envisioned. It almost goes without saying (except I'm saying it here), that it was only with the support of family, friends, collaborators, and mentors that I succeeded in finishing. The title page has only my name on it but their imprint is everywhere throughout the document.

I first would like to thank my committee co-chairs, Professors Erica Fuchs and David Hounshell. Erica and David, both in their own ways, exemplify the role of a thesis advisor. I am a better scholar and a better person for having had their guidance, both personal and intellectual, during this entire process. Next, I would like to thank the other members of my committee, Professors Granger Morgan and Atsushi Akera. Your insightful comments and questions forced me to consider the gaps in my research and helped me see through the end of this dissertation.

This research was made possible thanks to funding grants from the National Science Foundation Science of Science Policy Program (Award # 28935.1.1121844) and the National Institute for Standards and Technology (Award # 28994.1.1080278). I would also like to thank the National Science Foundation's Graduate Research Fellowship Program for its support.

Thank you to the Semiconductor Research Corporation for its support. Thank you to Larry Sumney, Ralph Cavin, Victor Zhirnov, Celia Merzbacher, and George Bourianoff for your time and feedback during several phases of this project. A special thanks to An Chen, Tom Theis, and Jeff Welser for their support as directors of the NRI in granting me access to the organization. Thank you to David Seiler who was an enthusiastic champion of this research and helped open doors along the way.

I am also grateful for the support from the entire staff of the Engineering and Public Policy department at Carnegie Mellon who were always willing to help me figure out what was needed and how best to get it done.

The list of friends and family who have seen me through this journey is too long to list here in totality. For your support and patience there is a debt of gratitude that I hope to repay with commensurate interest in the years to come.

Lastly to my parents, Amina and Naveed, and my brothers, Zain and Noor: we have all shared in this struggle and finishing would not have been possible without each of you.

Abstract

This dissertation is an historical and evaluative study of the semiconductor industry as it approaches the end of the silicon integrated-circuit paradigm. For nearly 60 years, semiconductor technology has been defined by rapid rates of progress and concomitant decreases in costs-per-function made possible by the extendibility of the silicon integrated-circuit. The stability of this technological paradigm has not only driven the transformation of the global economy but also deeply shaped scholars' understanding of technological change. This study addresses the nature of technological change at the end of a paradigm and examines the role and capability of different institutions in shaping directions and responding to challenges during this period.

This study first provides theoretical and historical context for the phenomenon under consideration. In order to place the dynamics of an industry at the end of a technology paradigm into proper context, particular attention is given to the semiconductor industry's history of failed proclamations of impending limits. The examination of previous episodes of technological uncertainty and the development of institutions to respond to those episodes is used to illustrate the industry's departure from previous modes of technological and institutional evolution.

The overall findings suggest that existing institutions may not be capable of addressing the semiconductor industry's looming technological discontinuity. Despite the creation of an entirely new institution, the Nanoelectronics Research Initiative, specifically oriented toward the end of Moore's Law the industry, government agencies, and the scientific community writ large have been unable to find a successor to the silicon CMOS transistor to date. At the terminus of this dissertation, research toward new computing technologies remains ongoing with considerable scientific, technological, and market uncertainty over future technology directions.

Table of Contents

List of Figures

List of Tables

Introduction

Technology change is at the heart of economic dynamism (Schumpeter 1942) and productivity growth (Solow 1957). Early theories of technological change from the neoclassical tradition differ on their emphasis on supply – scientific discoveries push new technology (Nelson 1962) – or demand – innovators responding to market demands (Schmookler 1962) – as the primary determinants. However, scholars note that these fail to address empirical regularities of the innovation process, such as consistently different rates of technology change by industry (Nelson and Winter, 1977). Furthermore, a new understanding of the nature of technology as artificat, knowledge, and process (Layton, 1974; Sahal, 1985; Vincenti, 1990) as opposed to simply applied science, precipitated a change in the understanding of technological evolution. Firms and engineers hold beliefs about the nature of technological progress and their beliefs are embodied and shaped by technological artifacts (Nelson and Winter, 1977; Sahal, 1985). In several industries, exemplar artifacts define the regular notion of progress (Sahal, 1985).

Two theories of technological change build on this conception of knowledge by borrowing from Kuhn's (1962) notion of scientific revolutions. Dosi (1982) proposes the concept of technology paradigms, akin to Kuhn's scientific paradigms, as focusing heuristics that serve to shape the search space of engineers and firms. Meanwhile, Constant (1980) proposes the concept of "presumptive anomaly" as a driver of technological revolution, which he views as a process embedded in the beliefs of a community of practitioners. Importantly, both theories place non-market institutions as central to the process of technological paradigm change (Dosi) or technological revolution (Constant).

This focus on non-market institutions is markedly different than research in the Schumpeterian tradition which has largely examined the market conditions for maximizing

innovation (see Cohen 2010 for a review) or the effect of innovations on existing firms (Tushman & Anderson 1986; Anderson & Tushman 1990; Henderson & Clark 1990). The literature examining the role of institutions in technology change has largely been focused on issues related to national competitiveness (Nelson 1993) such as that with the "systems of innovation" approach. Missing from both literatures is an examination of the ability of institutions to address the challenges posed by the end of a technology paradigm. This dissertation examines the semiconductor industry as it approaches the end of Moore's Law and possibly the end of the silicon integrated-circuit technological paradigm.

Gordon Moore's 1965 publication observing the exponential rate of growth in integrated circuit complexity came to be colloquially known as "Moore's Law." Since that time, scholars have argued that it is best understood as a self-fulfilling prophecy where the industry and its associated institutions actively worked to adhere to Moore's projections (Mollick 2006; Brock & Moore 2006; Schaller 2004). The 50-year Moore's Law trajectory and the semiconductor industry have been the subject of considerable scholarly research, and with good reason. Semiconductors are a prototypical example of what economists call "general-purpose technologies" (GPT) (Bresnahan & Trajtenberg 1995). Per Bresnahan and Trajtenberg (1995), GPT's are characterized by their "technological dynamism" and "pervasive use in a wide range of sectors." Many would regard these traits as self-evident for semiconductors where the rapid pace of advance along the Moore's Law trajectory has given rise to entirely new industries that have reshaped daily life for billions and simultaneously driven economic growth around the world (Jorgenson 2001; Jorgenson & Vu 2007). Semiconductors are also remarkable because the 50-year Moore's Law trajectory has occurred entirely within the silicon integrated-circuit platform. As a result, silicon integrated-circuits are an exemplar of a technological paradigm and

their evolution has greatly influenced the thinking of scholars that study technological change (see e.g. Dosi 1982).

Since the 1970s, the primary technical driver of semiconductor advancement, and thus Moore's Law, has been miniaturization of the metal-oxide-semiconductor field-effect-transistor (MOSFET). Despite a series of predictions about impending limits to progress in semiconductors which arose periodically as early as the 1960s and throughout the 1970s and 1980s, the miniaturization-driven density improvements largely continued unabated. However, beginning in the early 2000s, experts began to raise new warnings and predictions about impending limits to miniaturization and thus advancements in silicon integrated-circuits. In the years to follow, the nature of technical progress in the industry began to change as continued advancement was inhibited by materials limits. Most recently the industry and its associated collaborative institutions began to undergo rapid change as the Moore's Law trajectory faltered.

In order to continue apace with "Moore's Law" the industry must identify a new computing device, or switch, capable of replacing the MOSFET – which has prevailed as the dominant technology for over 30 years. The discovery and development of a new switch on a new (non-silicon) material platform requires investments in basic research, which today has largely moved out of the industry's corporate labs. A new switch in a new material system also threatens to de-value of the core competencies developed by firms across the supply chain in the industry – from designers and equipment producers to software programmers and semiconductor manufacturers themselves. At the time of this writing, there remains considerable scientific, technological, and market uncertainty about the future directions of the semiconductor industry. Given that reality, a particular focus of this dissertation will be examining and analyzing the

institutional response to date and whether it has been appropriate considering the scale of this technological discontinuity, i.e. the end of Moore's Law.

This dissertation has four main findings, each contained within a specific chapter as summarized below.

1. Existing scholarly literature has under-examined the dynamics at the end of a paradigm and when existing institutions may underperform in overcoming a technological discontinuity.
2. Over 50 years, the semiconductor industry's organizational, market, and institutional structures evolved around the Moore's Law trajectory. That structrue may no longer be capable of addressing technology needs as the industry's nature of technological progress changes.
3. Semiconductor industry leaders created a new institution, the Nanoelectronics Research Initiative, in response to a scientific conjecture about the limits of performance of emerging technologies. The NRI has been able to coordinate scientific research directions in emerging technologies.
4. Despite the presence of the NRI and related efforts, current institutions may not be enough to address the technological discontinuity. There are several implications of this case for theory and policy.

Institutions at the End of a Technology Paradigm

This chapter is organized as follows. First, I will provide a brief review of the broad field of the literature concerned with technology change. The purpose is not to provide a comprehensive review of the relevant literatures but to instead broadly outline the evolution of prevailing thought and research strategies through the present. Second, I will discuss in detail the concept of "technology paradigms." I will then discuss the shortcomings of the research that has followed in the tradition of the "technology paradigms" idea. I will focus on the challenges of studying the end of a technology paradigm and the prospects for paradigm transitions. Lastly, I will end this chapter by suggesting new avenues for research.

1. The Roots of Technology Change Studies

Schumpeter's writings laid the foundation for an ongoing examination on the relationship between the determinants of technical change and the impact of that change economic growth. Schumpeter called the process of "Creative Destruction…[,] the essential fact about capitalism" (Schumpeter 1942, p.83) and, seemingly in anticipation of the work of scholars to follow in his footsteps, described the varieties of this process while simultaneously challenging economic orthodoxy's primacy of price competition:

> But in capitalist reality as distinguished from its textbook picture, it is not that kind of competition which counts but the competition from the new commodity, the new technology, the new source of supply, the new type of organization (the largest-scale unit of control for instance)—competition which commands a decisive cost or quality advantage and which strikes not at the margins of the profits and the outputs of the existing firms but at their foundations and their very lives. (Schumpeter 1942, p.84).

Whereas in earlier work, Schumpeter saw the entrepreneur as the lifeblood of dynamic capitalism, in *Capitalism, Socialism and Democracy* Schumpeter took a different tact, arguing that "in the process of creative destruction, restrictive [i.e. monopolistic] practices may do much to steady the ship and to alleviate temporary difficulties" (Schumpeter 1942, p.87). Schumpeter further argued that once economists properly accounted for the dynamic effects of creative destruction, "the large-scale establishment…has come to be the most powerful engine of that progress and in particular of the long-run expansion of total output" (Schumpeter 1942, p.106). Furthermore, Schumpeter saw the ascendancy of the large, bureaucratically managed enterprise as endangering and rendering obsolete the individual entrepreneur's role in capitalism.

Solow's (1957) finding that 87.5% of the doubling in gross output per worker from 1909 to 1949 was attributable to technical change placed it squarely at the heart of economic growth. The NBER's 1960 conference and subsequent 1962 publication, "The Rate and Direction of Inventive Activity: Economic and Social Factors," offered a broad outline of the early study of technical change. In his introduction to the volume, Nelson identified four roots to the growing interest in the study of inventive activity: a series of publications on the nature of productivity growth [including (Solow 1957)], the Cold War and the importance of technical change and economic growth to military capabilities, a revisiting by economists of Schumpeterian ideas on the relationship between technical change and competition, and the postwar debate on science policy and subsequent creation of the National Science Foundation (Nelson 1962).

An early debate emerged within the community of scholars of technical change over the relative importance of market factors in directing inventive efforts. One view held that market signals "pulled" the effort of firms toward problems with unmet demand (Schmookler 1962). A second view held that the generation of new scientific knowledge opened up a search for

applications of that knowledge and eventually "pushed" technological discoveries into the marketplace (Nelson, 1962). The volume also addressed questions of the effects of institutional context on the direction of technical change. Arrow (1962) argued that the indivisibility, inappropriability, and uncertainty in the production of knowledge limited the incentive of private firms to invest in basic research, leading to underinvestment from a societal perspective. Other contributors highlighted the relative weakness of market factors in the decision-making process of research and development project selection by university researchers (Merrill 1962) and military leaders (Cherington et al. 1962). This early work presented a mixed and incomplete picture of technical change. During some periods, inventive activity seemed responsive to market factors, but these were unsatisfactory in explaining the host of new technologies that were closely linked to scientific advances. Furthermore, the evolving institutional context of the scientific enterprise in the postwar years seemingly influenced the direction of scientific research itself and possibly the supply of new scientific discoveries.

2. Neither Technology Push nor Demand-Pull

The early studies of technical change failed to explain adequately noted stylized facts about the process of innovation, such as vast differences in inter-industry rates of progress and the inherent uncertainty in the production of innovations, or account for the institutional complexity of the "selection environment" and its inter-industry differences (Nelson & Winter 1977). In addressing these shortfalls, researchers began to "open the black box" (Rosenberg 1982) of technical change by incorporating the nature of technology into their analyses. This approach represented a conceptual break from previous methods in which researchers largely relied on economic statistics to glean insights into the innovation process. Researchers now began to examine the process of technical change itself: why did certain technologies improve

faster than others (Nelson and Winter, 1977), why did firms (and engineers) choose to address the technical questions they did (Rosenberg 1969), and what was it about the technology itself that shaped these outcomes and decisions (Sahal 1985)?

The rest of this section will briefly discuss the evolving understanding of technology during this period. Following this discussion will be an in-depth examination of the ideas presented in Dosi (1982) as an archetypal example of how scholars incorporated the concept of technology to aid their analyses of technical change. This section also draws heavily on Constant (1980) as both a contrast and a complement to the ideas presented in (Dosi 1982a).

2.1. Technology as Knowledge, Process, and Artifact

Central to the conceptual advance of "opening the black box" was a change in how scholars viewed technology. Neoclassical growth theory represented technology as the output(s) of a particular production function but abstracted away from the of any product or process (Nelson & Winter 1982). The neoclassical approach to representing technology discussed above could be seen as an abstraction from the prevailing conception of technology as applied science as represented in processes (i.e. "how things are made") and artifacts (i.e. "what things are done or made") (Nelson & Winter 1982; Layton 1974). Scholars proposed a new expanded definition of technology as processes, artifacts, and knowledge (Layton 1974; Vincenti 1990). Layton (1974) argued this reconceptualization of the nature of technology had ramifications for the understanding of the processes by which technologies are shaped as well as the relationship between science and technology. Technical knowledge came to be understood as separate from but related to scientific knowledge and not simply applied science (Layton 1974; Vincenti 1990). This view meant that the study of how technical knowledge comes to be and evolves is worthy of

study, just as sociologists of science have studied the institutional, cultural, and organizational aspects of scientific knowledge (Kuhn 1962; Crane 1972).

This conception of technology also rendered it subject to the economics of information (Nelson 1959; Arrow 1962). There are important differences, however. Technical knowledge is indivisible, nonrivalrous in use but not costless to imitate, and costly to produce. Polanyi's (1958) work on tacit knowledge – "we know more than we can say" – influenced the thinking of researchers considering the production process. Using the analogy of a cake recipe, scholars noted that simply knowing all the steps and ingredients of a recipe did not guarantee a masterfully produced cake (Winter 2005). Similarly, firms possess technical knowledge that is greater than the sum of the tasks it employs in a production process (Nelson & Winter 1982). These insights have implications for understanding firm strategy and competitive outcomes.

2.2. Technological Paradigms and Technological Trajectories

Dosi begins his paper with a discussion of the insufficiency of existing "demand-pull" and "technology-push" theories. His argument rests on a definition of technology as "both directly 'practical' (related to concrete problems and devices) and 'theoretical' (but practically applicable although not necessarily already applied) know-how, methods, procedures, experience of successes and of course physical devices and equipment" (Dosi 1982a). Technology encompasses process, artifact, and knowledge. His next draws a parallel between the "'perception' of a limited set of possible technological alternatives and of notional future developments" and Kuhn's concept of "scientific paradigms." Per Dosi, a "technological paradigm embodies strong prescriptions on the *directions* of technical change to pursue and those to neglect," and these are based on technological knowledge, which he describes as

"selected principles derived from natural sciences and on selected material technologies" (Dosi, 1982).

Dosi's definition of technological paradigms is closely related to the idea of a "technological regime" (Nelson and Winter, 1977). Nelson and Winter describe a "technological regime" as "relating to technicians' beliefs about what is feasible or at least worth attempting" (Nelson and Winter, 1977). In both cases the beliefs held by individuals regarding technological possibilities influence the decisions of their respective firms. In his study of the turbojet revolution, Edward Constant explicitly includes community belief into his definition of technological knowledge, which he says, "comprises traditions of practice which are properties of communities of technological practitioners" (Constant 1980). Of the three theories, only Constant discusses the difficulty in defining the boundary of the technological community that subscribes to a specific outlook, noting that the hierarchical and non-exclusive nature of technologies makes these boundaries somewhat arbitrary, especially compared to scientific communities (Constant, 1980).

While each of those definitions shares the concept of community belief about technological possibilities, they differ in how these beliefs are embodied and how they shape future technology directions. Several scholars identified specific products (i.e. technological artifacts) as embodying the focusing device that defined the parameters of progress. In Nelson and Winter's discussion of technological regimes, they identify the "DC3 aircraft in the 1930's" as having "defined a particular technological regime" and note that "[f]or more than two decades innovation in aircraft design essentially involved better exploitation of this potential" (Nelson and Winter, 1977). For Nelson and Winter, the artifact defines the progress of technology by "focus[ing] the attention of engineers on certain directions" that they define as "trajectories and

strategies for technological advance" (Nelson and Winter, 1977). A related concept comes from Sahal (1985) in which archetypal examples of a technological artifact are "technology guideposts" which offer a "basic design" that is then subject to "bit-by-bit modification" along an "innovation avenue." Similarly, in discussing the evolution of competitive dynamics within an industry, Utterback and Abernathy (1975) use the term "dominant design" to describe a specific product that shifts the locus of innovation and competition within an industry away from product focused to process based.

Meanwhile, Dosi's paradigms are broader in scope and not strictly tied to a particular product or process. A technological paradigm is identified by the

> "generic tasks to which it is applied (e.g. amplifying and switching electrical signals), to the material technology it selects (e.g. semiconductors and more specifically silicon), to the physical/chemical properties it exploits (e.g. the "transistor effect" and "field effect of semiconductor materials), to the technological and economic dimensions and trade-offs it focuses upon (e.g. density of the circuits, speed, noise-immunity, dispersion, frequency range, unit costs, etc.)" (Dosi, 1982 page 153).

For Dosi, a paradigm defines what a technology does and how it does it and provides the rules for evaluating different substantiations of that technology. Within a given paradigm, a "trajectory" is the "pattern of 'normal' problem-solving activity (i.e. of 'progress')" (Dosi, 1982). Where problem solving occurs as engineers navigate the "multi-dimensional trade-offs among the technological variables which the paradigm defines as relevant" (Dosi, 1982). The lack of an explicit tie to an archetypal product also changes the scope of Dosi's trajectory vis-a-vis those of Nelson and Winter's trajectories or Sahal's innovation avenues. For Nelson and Winter (1977) progress along a trajectory shows that "innovation has a certain logic of its own," i.e. the logic embodied in the artifact. Dosi's trajectories allow for the same type of directed cumulative progress noted above but also encompass a wider range of technological possibilities

that incorporate feedback from market mechanisms such as changes in relative factor prices or demand. This responsiveness within a trajectory is akin to what Rosenberg (1976) termed "focusing devices" or Hughes' (1983) concept of "reverse salients," where engineers place effort on fixing the subsystem or component that has become a bottleneck to system-wide progress. Meanwhile, Constant does not specifically discuss the role of markets in shaping "normal technology," which he describes as "the improvement of the accepted tradition or its application under 'new or more stringent conditions'" (Constant, 1980).

The discussion so far has yet to address the fundamental question of how a technology paradigm emerges. Dosi provides only a very general discussion of his view of the forces that determine a paradigm. While he notes the "general weakness of market mechanisms in the ex-ante selection of technological directions especially at the initial stage of the history of an industry," his later explication clearly demonstrates that his mental model is one of firms establishing a technological paradigm through R&D processes that translate scientific discoveries into new technologies. As Dosi writes,

> "New technologies are selected through a complex interaction between some fundamental economic factors (search for new profit opportunities or for new markets, tendency toward cost saving and automation, etc.) together with powerful institutional factors (the interests and the structure of the existing firms, the effect of government agencies, etc.)" (Dosi, 1982 p 157)

Dosi later expands on the institutional factors by suggesting a role for "'bridging institutions between proper 'science' and technology" and "forms of institutional intervention which allow 'a hundred flowers to blossom'" (Dosi, 1982).

By contrast, for Constant the locus of technological change is community practice and has no explicit institutional or market component. Constant writes that

> "major discontinuities in technological practice occur within communities of practitioners, and the dynamics of that process of change can be better studied at the community level than at individual, firm, national, or industrial aggregate levels" (Constant, 1980 p 10).

Constant's model for the process of radical technological change, which he calls technological revolution, also hews more closely to Kuhn's model of scientific revolutions. Constant describes "presumptive anomaly" as occurring "when assumptions derived from science indicate either that under some future conditions the conventional system will fail (or function badly) or that a radically different system will do a much better job" (Constant, 1980 p 15). In response to the presumptive anomaly, an alternative technology is formulated and competition between old and new occurs via the adoption decisions of the afflicted community of practitioners (Constant, 1980). According to Constant's model, the new technology specifically challenges the community's existing tradition of testability as new technologies are not readily comparable along the dimensions the community uses to evaluate the existing technology (Constant 1980; Constant 1987).

Constant's "presumptive anomaly" may serve as a conceptual bridge between his theory of technological revolutions and Dosi's views on the emergence of technological paradigms. Dosi gives little treatment to the source of scientific discoveries that underpin a new paradigm, adding only that it is likely to be influenced by non-market institutions (i.e. government and universities). According to Constant the "presumptive anomaly" arises from within a community of practitioners, which spans organizational and institutional boundaries. Constant's model provides a motivation for directed search for new scientific solutions to as-of-yet un-encountered technological failures.

3. Neo-Schumpeterian Studies of Technical Change and Industry Evolution

Despite the acknowledged role of non-market actors in the process of technical change, much of the existing literature in the Schumpeterian tradition has centered around two strands of research, each of which places firms at its locus. The first examines Schumpeter's claims regarding the relationship between firm size, market power, and innovation. The second has attempted to unpack Schumpeter's "creative destruction" by understanding how innovations affect existing firms and influence the evolution of industries. This section will briefly review both of those literatures.

As discussed above, Schumpeter (1942) challenged orthodox views on the relationship between market power and innovation by arguing that a dynamic view of the process of creative destruction indicated that monopolies provided a more stable platform for long-range investment. Schumpeter (1942) was impressed by what he saw as the growing importance and centrality of the large, bureaucratically managed firm – at the expense of the entrepreneur – to technological progress. In scrutinizing Schumpeter's claims, economists have largely relied on firm and industry-level data to examine how firm size and market power affect investment in R&D. Gilbert's (2006) review of the theoretical and empirical literature examining the relationship between market structure and innovation noted that many of the findings are contingent on the particulars of a specific industry including the strength of the appropriability regime and whether an innovation is a product or process innovation. Cohen (2010) summarized the empirical findings on the relationship between firm size and research and development, finding that the propensity to conduct R&D increases with firm size and that larger firms tend to invest in more incremental and process innovation. Cohen's (2010) review also emphasizes the relative paucity of industry-level controls needed to understand how "technological opportunity and

appropriability conditions, as well as the degree of market segmentation," affect R&D investment by firms.

A second strand of literature has attempted to unpack the process of "creative destruction" by examining the relationship between existing firm capabilities and innovations. Paraphrasing Schumpeter, creative destruction is a kind of competition that strikes at the foundations of existing firms (Schumpeter, 1942). Scholars have identified different types of innovations that render obsolete existing firm competences. Abernathy and Clark (1985) argue that "radical" innovations, the type they link to creative destruction, are those that render obsolete existing technology and production competence but apply to existing markets. Tushman and Anderson (1986) argue that firms new to an industry are the most likely to supply the innovations that lead to technological discontinuities because existing firms innovate along their existing competencies. Henderson and Clark (1990) more explicitly link existing firm competency to the specific construct of its products, arguing that architectural innovations are those where the core design concepts remain the same but the linkages between components is changed and existing firms struggle to compete. Meanwhile, Christensen (1997) highlights technological innovation that disrupts firms' existing market knowledge by creating new markets or better addressing the needs of over-served customers through lower-end products. In each case, existing firms fail to compete with new innovations because they are unable to adjust their existing practices and processes to address the competitive challenge posed by the innovation.

A related literature has examined the link between the evolution of industry structure and innovative activity. Scholars have identified patterns of entry, shake-out, and exit of existing firms over the course of an industry's life cycle (Utterback & Abernathy 1975; Anderson & Tushman 1990; Utterback & Suárez 1993; Klepper 1997). According to one theory, the early

period of an industry is marked by entry of firms who compete via product innovations in a nascent market, and shake-out occurs when a "dominant design" emerges (Abernathy and Utterback, 1978; Utterback, 1979). With the emergence of a dominant design, products are standardized and the locus of innovation shifts toward process innovation and over time the most efficient and productive firms are the remaining survivors (Abernathy and Utterback, 1978). An alternative theory argues that the shift in the locus of innovation from product to process occurs because firms grow larger and larger firms are more capable of appropriating returns from process innovations (Cohen & Klepper 1996; Klepper 1997).

4. Clarifying the relationship between Products, Trajectories, and Paradigms

This section will briefly discuss further the concept of products, trajectories, and paradigms in order to clarify the relationships between the literatures reviewed in sections 3 and 4. Reconciling the gap between the theoretical literature and empirical findings will be one focus. The literature in section 3, which attempted to evaluate empirically the rate of technical progress, examined the progression of technology trajectories as expressed in product improvements (see e.g. Sahal, 1985). Similarly, much of the empirical literature examined in section 4 has focused on the relationship between competition and innovation in product markets. How do we reconcile the gap between the empirical literature's focus on products and the theories of technology change that emphasize paradigms as shared cognitive outlooks that operate beyond the bounds of any specific firm or product (Dosi, 1982; Constant, 1980)?

I will first review the relationship between changes in products, trajectories, and paradigms. Then I will use the semiconductor industry as an example to further clarify these concepts.

4.1. Paradigms, Trajectories, and Products

Firms compete in product markets. The products produced by firms have specific characteristics that improve over time, often along an observable trajectory (Nelson and Winter, 1982; Sahal, 1985). In industries with a stable trajectory and associated dominant design (Utterback and Abernathy, 1978), most of the innovations that occur are incremental and enhance the position of the already entrenched firms (Abernathy and Clark, 1985; Nelson and Winter, 1982; Dosi, 1982).

The theoretical literature in the evolutionary economics tradition often fails to account for the product market specificity of trajectories. Industries, broadly grouped by underlying technology, often contain several product classes with distinct but related trajectories.[1] To better apply the concept of trajectories in these cases, it may be useful to say products within an industry share a "core component" (Murmann & Frenken 2006). Core components, per Murmann and Frenken (2006), are components in a technology that affect a large number of that technology's characteristics. For example, civilian and military aircraft show different rates of progress but share the jet turbine as a core component (Frenken et al., 1999). Likewise, memory (DRAM) and logic (CPU) chips also have different rates of improvement in density but share the CMOS integrated circuit as their core component.

Understanding the product market specificity of trajectories is also important to understanding competitive dynamics within an industry. In the case of hard drives, Christensen and Rosenbloom (1995) demonstrated that drives made for different product markets (e.g. mainframes, desktops, and laptops) exhibited different improvement trajectories, with drives

[1] For an example, see Klepper and Thompson (2006) on sub-markets in the laser industry.

developed for new markets improving faster than those in existing markets. As a result of the faster rate of progress, firms that led the way in introducing drives for new markets often emerged as winners as they captured market share from firms that lagged in introducing these architectural innovations (Christen and Rosenbloom, 1995; Christensen, 1997).

The ability of architectural innovations to increase the slope of a trajectory also calls for more research on understanding the primary driver behind a trajectory's slope. The notion of trajectories described above makes implicit the idea that the primary driver of progress within a product category is something akin to the Murmann and Frenken's (2006) core component. However, in the case of hard drives (Christensen 1992a; Christensen 1992b) and optical lithography (Henderson 1995), engineers were able to extend the life of incumbent technologies well beyond the espoused limit inherent to the core component. One possible explanation is that existing trajectories tend to focus the effort of engineers on methods of problem solving, and as a trajectory and associated component matures, the space of possible solutions examined by the community of practitioners expands particularly as those invested in the existing ecosystem attempt to preserve the value of their investments (Adner & Kapoor 2016).

Unlike products and trajectories, which tend to have observable characteristics, paradigms consist of both an artifact and a socio-cognitive aspect. Nelson and Winter (1982) described these as the "beliefs" of engineers and technicians. Dosi (1982) suggests that a paradigm defined an "outlook" and a "set of heuristics" that both dictate the embodied in the variables and parameters by which a technology is evaluated by engineers. Dosi's discussion of the physical representation of a paradigm is largely abstract, noting only that a paradigm is identified by the "generic task…[,] materials technology…[,] and physical/chemical properties" it exploits (Dosi, 1982). Constant (1980, 1987) offers a link between the role of the artifact and

community belief. For Constant (1987) a community's "tradition of technological practice" embodies "higher-level traditions of technological testability, which in turn is composed of both some set of testing technologies and techniques and some set of normative values." Constant (1980) argues that because "[t]echnological systems directly, not vicariously, explore the environment," the acceptance of a radical technology rests on some proof of concept. In other words, technologies must work in the real world. A paradigm cannot be built around a technology that lacks some proof of concept, even if that proof of concept is unrefined.

The last remaining relationship to discuss is that between a paradigm and the trajectories along which it focuses the evolution of a technology. As noted above, a trajectory is measured based on the characteristics of products. This requires a product built around or inspired by the paradigm-defining artifact to become the start of a trajectory. Implicit in Dosi's notion of paradigms and trajectories is the idea of a 1-to-1identification between the two; i.e., a paradigm has a single dominant trajectory. However, as argued above, products within an industry display related but distinct trajectories. It may be more exact to argue that there exists a paradigm-wide trajectory that is shared amongst product classes. Furthermore, there exists the possibility for multiple, distinct trajectories within a single paradigm. The notion of successive trajectories is akin to Foster's (1986) S-curve theory where a specific technology reaches its limits and is supplanted by another. Once again, delineating between a shift in trajectory versus a shift in paradigms requires understanding the hierarchical nature of a specific technology.

4.2. An Example from the Semiconductor Industry

In order to bring clarity to and further crystallize the ideas presented in the previous section, I will further discuss the relationship between products, trajectories, and paradigms using the example of the semiconductor industry.

Since the 1960s, the dominant paradigm of the semiconductor industry has been that of the *silicon integrated circuit.* However, the materials technology (silicon) and functional form (integrated circuits) that constitute the core of the paradigm were not commercialized until almost a decade and a half after the invention of the transistor in 1947. The earliest transistors were fabricated using germanium; silicon transistors were first commercialized in 1954 by Texas Instruments. The integrated circuit emerged as a solution to the "tyranny of numbers" (the problem of ensuring circuit quality and ruggedness with exponentially expanding numbers of discrete components in a circuit) but was not one of the approaches funded by military agencies during the late 1950s. Instead, the invention of the integrated circuit relied on advances in diffusion and silicon dioxide processing, first achieved at Bell Labs during the 1950s.

The first shift in technology trajectory in the semiconductor industry came by way of the integrated circuit. Prior to the invention of the integrated circuit, firms focused on improving the characteristics of individual transistors. As Gordon Moore (1965) argued, however, the prospect of continually increasing device density of integrated circuits opened up new technological capabilities. At the time of his writing, the commercial market for silicon integrated circuits was vastly overshadowed by government (i.e. military) purchases, and the market for transistors would continue to be larger than that for integrated circuits through the 1960s. However, the success of this new trajectory was dependent on two other innovations: a new manufacturing technology better suited to maximizing density (MOSFET) and a product that benefited from increased density (semiconductor memories, specifically DRAM).

During the 1960s a new transistor technology, the metal-oxide-semiconductor field-effect-transistor (MOSFET) was reduced to practice by industry researchers with significant contributions coming from Bell Labs and RCA. This new device technology was significantly

slower than existing bipolar-junction transistors but offered simpler manufacturing processes and greater density. Importantly, MOS technology also exhibited preferential scaling characteristics. As first codified by Dennard et al. (1974), reducing the size and operating voltage of a MOS transistor enabled higher densities at faster speeds with lower overall power consumption. MOS-based ICs first found applications in devices such as electronic watches and calculators, where speed was not essential. The first product that fully utilized the higher-density capabilities of the MOS-based IC and offered the promise of continued expansion along Moore's predicted trajectory was the dynamic random access memory (DRAM), which was first available commercially in 1970. Despite Moore's (1965) elucidation of a trajectory, the market for increasingly dense devices had yet to materialize prior to the widespread adoption of semiconductor memories.

The following year Intel also introduced the first microprocessor (or CPU), the 4004. Together, CPUs and DRAMs have remained the two single largest products in the semiconductor industry. Their rapid rate of advance opened up entirely new markets for personal computing devices, and the stability of the Von Neumann paradigm in computing--as well as the dominance of the Wintel duopoly--helped keep these two products the single largest products in the industry through the 1990s and 2000s. It is these two product classes – DRAM and CPUs – that have roughly followed Moore's Law since the 1970s.[2] For example, beginning in 1994, the semiconductor industry's roadmap projections were based on DRAM and CPUs and focused on solving technical issues for continued scaling of MOS manufacturing technology. By the 1980s, MOS-based integrated circuits became the majority of all integrated circuits by total sales, and

[2] In his 1975 conference presentation, Moore's graphs do show bipolar and MOS ICs following a similar trajectory over the 10 years since his 1965 publication (Moore, 1975).

by the mid-90s over 95% of all integrated circuit sales were CMOS – complementary MOS, a particularly energy efficient form of MOS devices.

However, despite the strength of the Moore's Law trajectory for MOS-based ICs, alternative trajectories persisted in other product categories. As one example, bipolar logic, which continued to be used in mainframes through the early 1990s due to its superior speed capabilities, did not follow the Moore's Law trajectory (Bassett 2002). IBM's 3090 mainframe, released in 1985, actually had lower circuit density than the IBM 3081, which was released in 1980; performance improvements in bipolar logic were achieved primarily through advances in architecture, circuit design, and packaging (Bassett, 2002). Other product categories, such as analog integrated circuits, also did not follow the Moore's Law trajectory. Yet, many of these other product categories also benefited from the rapid progress exhibited by the Moore's Law trajectory. The growing markets and fierce competition in the DRAM and CPU markets during the 1980s and 1990s drove investment in scientific research and manufacturing technologies that have been more widely applicable to products across the industry.

Silicon integrated circuits and "Moore's Law" are perhaps the canonical example of a technology paradigm and its associated trajectory. However, as the abbreviated history above shows, the paradigm emerged prior to any successful demonstration of the trajectory, and the trajectory was limited to a subset of the industry's products. The materials technology (silicon, 1954) was dominant far before the form factor (integrated circuits, 1961). Additionally, even after the emergence of the silicon integrated circuit, the emergence of the industry's dominant trajectory – i.e., Moore's Law – rested on the commercialization of a new manufacturing technology (MOS transistors) and a product that benefited from the rapid increases in density (DRAM). Furthermore, the trajectory was confined to products that utilized MOS manufacturing

technology and primarily driven by two product classes: DRAMs and CPUs. However, despite the localization of the trajectory to only a few major product classes, the entire silicon integrated circuit paradigm benefited from the rapid progress in scientific understanding and process technologies made possible by Moore's Law.

5. Research for Understanding Technology Paradigm Shifts

This section will discuss why paradigm changes are different than other examples of technical change. In particular, it will focus on the role of non-market based institutions in influencing paradigm change. Then, I will review the existing literature on the role of institutions in shaping paradigm changes and end with a suggestions for future avenues of research and open questions which remain.

5.1. What differentiates paradigm changes?

Paradigm shifts are unique examples of technical change for several reasons. First, the notion of a paradigm change is dependent on a given technology's hierarchical structure. A paradigm shift in a particular component may be modular or incremental with regards to the entire system (Constant, 1980; Vincenti, 1990). For the purposes of this discussion, a paradigm shift refers to a radical change in the nature of operation for a core component (Murmann and Frenken, 2006) in an open technological system (Tushman & Rosenkopf 1992). Murmann and Frenken (2006) explain the rarity of substitutions of a core component by borrowing from a biological analogy in which individual mutations are unlikely to increase overall fitness; thus mutations to genes that affect a large number of traits are increasingly unlikely to persist in the population.

Existing research offers numerous explanations for the low probability of success. Murmann and Frenken (2006) suggest that the low probability of success is due to myriad architectural and component changes needed throughout the system to achieve higher overall performance with a new core component. Evidence from failed, or delayed, technological transitions in both the hard disk drive and photolithography industries suggests that adoption of new technologies is mediated by the existing value network (Christensen and Rosenbloom, 1995; Henderson, 1995; Adner and Kapoor, 2016). Dosi (1982) suggests that a new paradigm relies on the availability of new scientific knowledge, and any technologies built on this new science must undergo selection processes that are mediated by a host of non-market institutions. Constant (1980) argues that the impetus for radical technologies comes from presumptive anomaly, which is rooted in scientific conjecture about the future failure of existing technology, and the eventual development of an entirely new tradition of practice by a new technological community for the new technology. Thus existing research suggests that technological paradigm changes involve actors from a variety of market and non-market institutions resolving questions regarding scientific, technological, and economic uncertainty.

5.2. Institutions at the end of a paradigm

Previous research suggests that complex market and technology dynamics during periods of paradigm shift indicate a role for myriad actors across different organizational and institutional contexts. This section reviews the extant literature on institutions at the end of a technology paradigm. I divide these institutions into three groups: markets and market actors, governments, and cooperative institutions.

5.2.1. Markets and Market Actors

The tension in the literature over which firms are likely to introduce innovations can be traced back to Schumpeter himself. Schumpeter "Mark I" (1911) argued for the primacy of the entrepreneur, while Schumpeter "Mark II" (1942) was astounded by the capabilities of large firms to continue delivering systematic improvements in their products and even entirely new technologies.

Understanding the ability of firms to overcome the scientific, technological, and economic hurdles at the end of a paradigm may require an understanding of the locus of technological progress within an industry, an understanding of the industry structure and extent of integration, and of the hierarchical structure and extent of modularization of the technology (Constant, 1987; Tushman and Rosenkopf, 1992). Scholars have identified industry differences in the role of internal and external research efforts and appropriability mechanisms (Pavitt 1984; Levin et al. 1987; Cohen et al. 2002; Cohen & Levinthal 1990). The literature on product (Utterback and Abernathy, 1985) and industry life cycles (Klepper, 1996) does not explicitly address the question of vertical disintegration. However, it is likely that industries with dominant designs begin to develop stable design rules, giving rise to modularity (Baldwin & Clark 2000). In these cases, as the industry structure evolves alongside the increasingly modular technology structure, firms may need to maintain system-wide knowledge in order to overcome architectural shifts (Chesbrough & Kusunoki 2001; H. Chesbrough 2003). In the case of the semiconductor industry, research suggests that despite widespread vertical disintegration the few remaining vertically integrated manufacturers maintained competitive advantage by focusing on delivering systemic innovations (Kapoor 2013). However, researchers have found that incumbent firms struggle to adapt to innovations that disrupt their existing market or product knowledge, and thus

entrant firms are likely to introduce and capture the value of these innovations (Tushman and Anderson, 1986; Henderson and Clark, 1990; Christensen, 1997).

An open question is whether the changing nature of industrial research, as recent evidence indicates that firms have reduced investments in basic science, even as overall spending on R&D has increased (Arora et al. 2015), affects the ability of firms to integrate scientific discoveries into radically new products.

5.2.2. Governments and Public Actors

Despite the historical examples of, e.g., the transistor, computer, and internet (National Research Council 1999) and the acknowledged role for government in long-term technology development where uncertainties are high and the fundamental underlying science is unknown (Nelson 2004), there is a gap in the literature on the potential role of government, if any, in orienting institutions to address technological discontinuities. Previous research has argued that the institutions designed to guide science and technology are products of national competitiveness efforts (Nelson 1993) and, in the case of the U.S., the result of political compromise (Hart, 2010). Thus, concentrated public efforts in technology have primarily occurred for the purposes of defense (e.g. the Manhattan Project) or technology catch-up in industries deemed crucial (e.g. SEMATECH) to national security and economic welfare. International examples are similar. British "research associations" were established late in World War I to try to bring British industry up to the research standards of Germany and the United States (Mowery and Rosenberg, 1991), and Japan's Ministry of International Trade and Industry (MITI) guided the restructuring of the Japanese computer and semiconductor industries to be

more competitive with US firms with the use of publicly subsidized cooperative research programs (Sakakibara, 1993; Ouchi, 1984; Flamm, 1988; Sigurdson, 2004).

Additionally, government institutions may influence the rate of scientific discovery and the viability of different appropriability strategies. The conjectured importance of scientific discovery to paradigm shifts in Dosi (1982) and noted growing importance of publicly funded science to firm patenting (Narin et al., 1997) suggests a role for public funding. Unexplored is how firms facing paradigm shifts utilize public scientific research if the science required to address a paradigm is outside a firm's traditional areas of expertise or how government agencies may best direct funding toward scientific research relevant to addressing challenges at the end of a paradigm. Another important role of the government in shaping technology directions at the end of a paradigm is the effect of policy on appropriability mechanisms, particularly patents. Given the need for new science and the importance of universities in generating that science, the effect of the change in patenting strategy by US universities in recent decades on technology transfer in emerging technologies remains uncertain (Nelson 2004; Mowery 2011).

5.2.3. Cooperative Technical Organizations

A growing literature has widened the locus of innovation beyond the firm to cooperative institutions that shape technology (Constant 1980; Rosenberg & Nelson 1994; Powell & Grodal 2005; Freeman 1991). Existing research has examined the role conferences (Garud & Rappa 1994; Garud 2008), communities of experts (Haas 1989; Rosenkopf & Tushman 1998), platform leaders (Gawer & Cusumano 2002), and embedded network agents (Fuchs 2010) can play in technology direction-setting during periods of technological uncertainty. Within the context of the semiconductor industry, research has found that industry roadmaps served to coordinate the

efforts of firms, academic researchers, and consortia along the industry's established trajectory (Schaller 2004) and that the nature of collaborations changes over a technology's life cycle with consortia playing a more central role during periods of emergence, when technologies have higher scientific uncertainty (Kapoor & McGrath 2014).

Of the papers cited above, only a handful explicitly focus on the role of cooperative institutions in addressing technological discontinuities. The technologies they examine include cochlear implants (Garud and Rappa, 1994; Garud, 2008), flight simulators (Rosenkopf and Tushman, 1998) and turbojet engines (Constant, 1980). In the case of cochlear implants and flight simulators, the technology of interest is a closed, assembled system (Tushman and Rosenkopf, 1992), and the technological discontinuity results from the decision of relevant regulatory bodies. As a result, regulatory agencies also play a pivotal role in the cooperative institutions. The conferences that settle scientific uncertainty about the safety of competing technology approaches in cochlear implants were sponsored by the NIH (Garud, 2008). Similarly, many of the new members of the cooperative technical organizations (CTOs) founded during the era of ferment in flight simulators were regulators (Rosenkopf and Tushman, 1998). Rosenkopf and Tushman's (1998) examination of CTOs in the flight simulator industry did not discuss how individual membership in different CTOs influenced technology directions. In the case of cochlear implants (Garud and Rappa, 1994; Garud, 2008), scientists and firms used conferences to standardize metrics and establish evaluation routines in order to compare competing approaches (i.e. single and multi-channel cochlear implants). Meanwhile, Constant's (1980) study of the turbojet revolution describes the process of community formation around a new technological paradigm. Importantly, although the scientific conjectures regarding the viability of turbojets arose from industry outsiders, the market for the new technology was

dominated by existing firms, in part due to the support of government R&D agencies (Constant, 1980).

5.3. Questions for Future Research on Paradigm Changes

The current understanding of the process of technological paradigm shifts remains in its early stages. Advancing the study of technological paradigm changes beyond abstract theory (e.g. Dosi, 1982) or historical cases (Constant, 1980) and toward a fleshed out taxonomy is an ambitious undertaking. Previous researchers have suggested that these outcomes depend on the competence of existing organizations (Abernathy and Clark, 1985; Tushman and Anderson, 1986), the structure of the technological system (Constant, 1980; Hughes, 1983; Tushman and Rosenkopf, 1992; Murmann and Frenken, 2006), and the relevance of non-market institutions such as external scientific organizations or appropriability mechanisms (Dosi, 1982; Levin et al., 1987; Cohen and Levinthal, 1990; Cohen et al., 2002). Scholars must also contend with changes to the institutional structure of innovation systems. While much of the early literature in the field was built around the mental model of a vertically integrated firm with a dedicated corporate research organization, firms have changed their approach to industrial research. Additionally, the organization of public science has been affected by the changing structure of research funding and changes in university patenting behavior as a result of policy changes (Nelson, 2004; Mowery, 2011). Developing a taxonomy of paradigm changes will require addressing open questions regarding the proper role and structure of non-market institutions in technological paradigm change and the effect of different technological and market conditions on the survival of existing firms.

Chicken Little and the End of Moore's Law

1. Introduction

As in the well-known children's book, *Chicken Little*, in which the main character cries "the sky is falling" one too many times, imperiling the other barnyard animals who have elected to ignore Chicken Little's latest dire warning, experts in the U.S. semiconductor industry have, almost since the industry's inception in the middle of the last century, sounded periodic alarms about the impending demise of the industry. As with the barnyard animals in *Chicken Little*, industry and government representatives have gotten stirred up by these alarms, only later to learn that the end of the industry had not come. Today, however, a silicon version of Chicken Little's last alarm is sounding, "The end of Moore's Law is near; the end of Moore's Law is near." As in *Chicken Little*, those parties who previously stirred in response to earlier warnings about the demise of the U.S. semiconductor industry are mostly shrugging their collective shoulders or have tuned it out entirely. As in *Chicken Little*, however, this time the danger is clear and present, and the consequences to American society – and the world – of ignoring what is happening are almost beyond calculation. The economic gains – specifically the gains to the U.S.'s and the world's productivity attributable to sustained increases in performance and lowering of costs of integrated circuits – have been clearly demonstrated. The social benefits of advancing computing at ever-smaller size and -lower cost surround us, with many visions of the future dependent on this trend continuing yet to be realized. Nature's limits to feature size of conductors and transistors in ICs, however, are upon us. This paper employs diverse methods to tell the silicon-equivalent of *Chicken Little*. It concludes with an assessment of the current situation faced by the semiconductor industry, the U.S., and the world and makes

recommendations for a response proportional to the real threat posed by the end of Moore's Law and the current state of the global semiconductor industry.

2. Chicken Little

There is a well-know children's story dating back well before the Brothers Grimm and titled variously, from *Henny Penny* to *The Sky is Falling*. The story endures in often less grim retellings, known most often as *Chicken Little*. The first American edition of the story, *The Remarkable Story of Chicken Little* (1840) opens, "Did you ever hear of Chicken Little, how she disturbed a whole neighborhood by her foolish alarm?" To make a short story even shorter, typically a leaf, acorn, or something else trivial falls on Chicken Little's tail, often when she—Chicken Little is almost always female—is grazing somewhere or doing something that's not quite right in the barnyard. Alarmed by the unknown sensation on her tail, Chicken Little runs around the barnyard screaming, "the sky is falling," getting all the animals panicked and beating hasty retreats to their respective safe houses. After a while, the panic dissipates when members of the community see that the sky is still where it's always been. After several reprises of Chicken Little's "the-sky-is -falling!" alarums, the barnyard creatures begin to ignore her—to keep doing what they're doing rather than move to a safe place. Then one day, there really was a major crisis at the farm (the appearance of a sly fox in most tellings), first noted by Little Chicken. But her cries are ignored. Despite Chicken Little's warning, the animals have been lulled into complacency by their prior observations that the sky's not falling—by Chicken Little's false alarms. Consequently, in many versions, Chicken Little and many other barnyard beasts meet their demise in the fox's jaws.

The dominant lesson of *Chicken Little* for children is intended to be, of course, "false alarms are bad for the community, so don't make them." The idiom "crying wolf," from Aesop's Fable, embodies this same lesson. As with most parables and "fairy tales," however, underlying questions and doubts can set in on the part of the listener or reader, including the precocious child: "What was really going on here? Despite the sky's not really falling, weren't some of Chicken Little's early alarums rooted in some basic phenomenon that was not well understood? What responsibility to Chicken Little and her concerns did the community have other than blowing off those concerns—and Chicken Little?" For, in many of accounts of Chicken Little, not just Chicken Little but many of her community's finest met their demise. "Why? Who's really at fault here?"

What follows is an analytical story about systematic performance improvements and cost reductions in semiconductor devices, known endearingly as *Moore's Law* after industry pioneer Gordon Moore. It is also a story of periodic alarms by members of the semiconductor community about imminent limits to Moore's Law. It is, more or less, a 21st-Century adaptation of *Chicken Little*. Chicken Little is played not by a young hen but several industry experts and observers who appear periodically to declare that they see the end of Moore's Law. The sky is played by the semiconductor industry's growth trajectory, the universe of firms and suppliers that make up the industry and provide jobs to tens of thousands, and the larger economy in which the industry exists and whose real growth has for decades been driven by the industry. The barnyard animals are made up not only by Chicken Little but other industry engineers and scientists, academic researchers, stockholders in the industry, government policymakers (including defense technology leaders), and semiconductor users writ large, from personal computer manufacturers

to specialty electronics makers and from automakers to the "average consumer" of cars, mobile phones, TVs, and video games.

3. Before Moore's Law

The research of economists such as the John Bates Clark Award winning Harvard Professor Dale Jorgenson suggests that the systematic declining of costs and increasing of performance of semiconductors since the innovation of the integrated circuit in the early 1960s (a.k.a., "Moore's Law") was the "foundation for the American growth resurgence" during the 1990s (Jorgenson 2001). Since the 1990s investments in information technology were the leading source of growth worldwide (Jorgenson & Vu 2007). Whether in the global economic North or South, the "chip" has been responsible for geometric decreases in the cost and the space in which that computation can be achieved with concomitant increases in computational performance, hence increasing technology adoption and usage. This phenomenon has led to relatively easy-to-measure increases in national and international productivity. Moreover, seemingly continuous innovation in semiconductors has bred innovation in other new technologies that have in turn contributed to economic growth and welfare. Cheap, small, fast, and powerful chips underlie a welter of productivity-enhancing technologies that increase quality of life in a way that is far harder for economists to measure. In short, the sky of the semiconductor story has been growing, kaleidoscopic even as the "size" of the world seemingly shrinks with the chips. This sky is certainly worthy of being a main character in a 21st-Century update of Chicken Little.

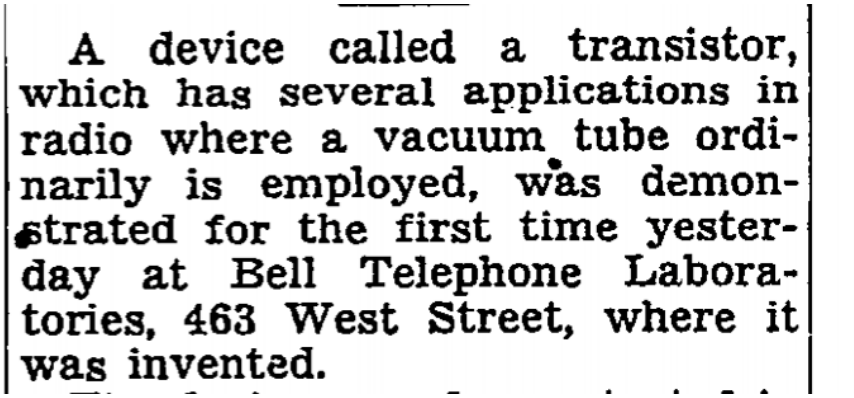
A device called a transistor, which has several applications in radio where a vacuum tube ordinarily is employed, was demonstrated for the first time yesterday at Bell Telephone Laboratories, 463 West Street, where it was invented.

Figure 1 - New York Times coverage of transistor demonstration. Page 46 of the July 1st, 1948 edition in a section titled, "The News of Radio."

Ironically, this great sky of semiconductor technology and the semiconductor industry is based on the 1947 invention of a technology so small in size, feeble in initial performance, and balky in operation that it barely caught the notice of *The New York Times (see Figure 1)*—the point-contact transistor. The point-contact transistor was the product of a fundamental research program initiated in 1945 by Mervin Kelly, the research director of Bell Telephone Laboratories, the research and engineering arm of AT&T/Bell Telephone, the company that enjoyed a sanctioned monopoly in the U.S. from 1913 until its breakup in 1983. Though it started small in both physical size and performance, the transistor came to be regarded as one of the 20th-Century's and Second Millennium's greatest inventions. The tiny transistor's vaunted position in the history of technology occurred thanks to its subsequent incorporation into the integrated circuit (IC), invented in the late 1950s. Although ICs were initially extremely expensive and used only in Cold War military technologies such as Intercontinental Ballistic Missiles, they would eventually take over nearly all electronic circuits, be they vacuum-tube-based circuits or printed "microcircuits" of discrete components. Gordon Moore made his earliest observations about the growth of the number of transistors in state-of-the-art ICs (Moore 1965) well before ICs had pushed out these other circuit types and technologies and well before his eventual

company's (Intel's) microprocessors came to dominate the insides of personal computers. Moreover, well before Moore delivered his now-famous paper that would become transmogrified into its eponymous Law, the first Chicken Little had made his[3] appearance in the barnyard. Let the story of the silicon industry and Chicken Little begin!

3.1. Experimentation, Variation, and Emergence of the Extendibility Paradigm

Although Bell Labs' point-contact transistor found its way into a handful of early applications—mostly within the Bell system itself—it ultimately proved to be too unstable and difficult to manufacture to be widely commercialized. In its stead the bipolar-junction transistor (BJT), conceived and patented by Bell Lab's Shockley in 1948, emerged as the first transistor broadly adopted and manufactured by firms in the emergent semiconductor industry. The rapid diffusion of the transistor in the 1950s stemmed from an agreement reached between AT&T and the Antitrust Division of U.S. Department of Justice (with the Pentagon's blessing) that forbade the telephone company to manufacture transistors for applications outside its own system. The agreement required the company not only to openly license its transistor patents but also to hold a three-part series of conferences, beginning in 1951, that ended with licensees' acquisition of what was colloquially called "Ma Bell's Cookbook." This cookbook contained papers, patents, and descriptions of what Bell Labs and its co-owner Western Electric Manufacturing Company knew about transistors.[4] These conferences served to diffuse knowledge about the fundamental

[3] In our story, all the Chicken Littles of silicon are male.

[4] Of course, not "everything" that Bell researchers and engineers knew about semiconductors and transistors could be embodied in words, pictures, and diagrams, because a great deal of their knowledge was what Michael Polanyi termed "tacit knowledge." The Bell Labs transistor conferences served to impart the organization's "formal knowledge" and some, but not all, of its tacit knowledge about semiconductors and transistor design, operation, manufacture, and application. Movement of Bell Labs researchers to other firms was also instrumental to the rapid diffusion of the transistor.

science of semiconductors, transistor design and operation, and modes of transistor application. Military contractors and their researchers were particularly prominent among the dozens of initial transistor licensees. Consequently, based on Bell Labs' core theoretical advancements, many subsequent contributions to the art of semiconductor manufacturing would come from both large, established enterprises and small, young firms.

The diffusion of knowledge was but one of several roles the U.S. military played in shaping the development of semiconductor technology and the industry during this formative period of transistor development. Military agencies provided direct funding for R&D and production facilities, served as knowledge clearing houses for disseminating best practices between firms, and defined technology needs and trajectories as an early, eager, and very large customer. The stringent requirements of military applications—e.g., miniaturization of circuits, low power consumption, and high reliability in rugged, high-temperature environments—served to elevate silicon over germanium as the industry's material of choice. Although germanium's higher electron mobility provided for faster speeds in early transistors, its sensitivity to elevated temperatures made it unsuitable for deployment in military electronics.[5]

The bipolar-junction transistor's path to commercialization highlights the importance of materials and process capabilities in enabling the industry's technological trajectory. Crude processing techniques limited the viability of the earliest implementations of the BJT. Both the grown-junction (Bell Labs, 1951) and the alloy-junction (GE and RCA, 1951) transistor were roughly an order of magnitude (i.e., 10x) slower than point-contact transistors owing to poor control of the device's doped regions. Advances in crystal growth and refinement techniques for

[5] During this period, national security needs also shaped the development of computer architecture and design. The first general-purpose transistorized computer was likely the "SOLO" built for the NSA by Philco and installed between 1956 and 1958. (Ceruzzi 2003)

both germanium and silicon enabled the development of new diffusion techniques for dopants. These innovations delivered an order of magnitude improvement in the operating speeds of silicon BJTs, putting them on a par with the fastest point-contact transistors. Additionally, silicon transistors—first commercialized by Texas Instruments in 1954—brought improvements in reliability crucial for the defense market. By early 1955, researchers at Bell Labs had committed to the diffused-junction silicon bipolar-junction transistor as its dominant design (Riordan & Hoddeson 1997 pp 223).

Even as the silicon BJT appeared to many to be a "stable" technology—or dominant design—the technology was maturing rapidly. Some researchers—call them the first Chicken Littles of silicon—projected definite limits to progress in BJT transistors and their applications, at best two orders of magnitude (i.e., 10^2 or 100x) away (Early 1959; Goldey & Ryder 1963; Wallmark & Marcus 1962; Johnson 1965). One common but now clearly erroneous assumption in each of these predictions was a fabrication limit of no smaller than a micron. Writing in 1962, RCA's J.T. Wallmark and S.M. Marcus concluded that the "minimum size of active components in a computer is within a factor of 2-5 [i.e., 2 to 5] of the dimensions of the active region of many devices of today," limited by variation in doping, lithography tolerances, power density, and cosmic ray bombardment. The tunnel diode, invented in 1957, was widely predicted to be a transistor killer due to its faster speeds. For example, an entire session of the 1960 Solid-State Conference was devoted to tunnel diode circuits. Thus, predictions of the end of transistor scaling and the search for alternatives to silicon transistors preceded the rosy implications of what Gordon Moore observed in 1965 in his now-famous paper, "Cramming more components onto integrated circuits" (Moore 1965).

While many of the foreseen device limits proved to be erroneous, by the end of the 1950s transistors faced pressing reliability challenges at the circuit level. Although transistors offered advantages in power consumption and waste heat relative to vacuum tubes, they faced similar integration challenges. Operators were still hand-soldering connections between discrete devices, and these imperfect connections often led to circuit failures. The "tyranny of numbers"[6] imposed serious reliability problems as electronic circuitry grew increasingly complex. Reliability problems especially concerned military clients, and leading military agencies in the late 1950s and early 1960s began to fund diverse approaches to reliable integration. Among these approaches, "micro-modules" appeared to be a conservative pathway reliant on existing processes, while "hybrid circuits" combined thin-film approaches for passive circuit elements with conventional semiconductor devices to emphasize ease of manufacture, low-cost, and reliability.[7] Based on a fundamental paper by MIT's Arthur R. von Hippel ["Molecular Engineering" *Science* 123, 3191 (1956): 315-317], Westinghouse Electric and the Air Force pursued "molecular electronics," an altogether radical re-imagining of semiconductor design that promised to yield miniature, integrated devices (Holbrook 1995; Brock & Laws 2012)[8]. However, Westinghouse's vision for molecular electronics remained vague and ill-defined.[9]

[6] Writing in 1958, Bell Labs' Jack Morton (1958) offered the "tyranny of numbers" as a primary explanation for the slow adoption of electronics technology in new fields.

[7] By 1969, open market sales of integrated circuits outpaced those for hybrid circuits by 4x (ICE 1971) but in-house production of hybrid circuits, primarily by IBM, remained significant through the 1970s.

[8] David C. Brock and David A. Laws, "The Early History of Microcircuitry: An Overview," *IEEE Annals of the History of Computing* 34, 1 (2012): 7-19. See also Edgar A. Sack and David A. Laws, "Westinghouse: Microcircuit Pioneer from Molecular Electronics to ICs," *IEEE Annals of the History of Computing* 34, 1 (2012): 74-82 and Hungsub Choi and Cyrus C. M. Mody, "The Long History of Molecular Electronics: Microelectronics Origins of Nanotechnology, "Social *Studies of Science* 39, 1, (2009) 11-50.

[9] By 1962 Westinghouse had begun to use the term molecular electronics interchangeably with monolithic integrated circuits (Holbrook, 1995).

By the middle of the 1960s, the diversity of approaches to miniaturization gave way to a new dominant design, the integrated circuit. Rather than producing hundreds of transistors on a wafer, slicing the wafer to separate and test the individual transistors, and finally painstakingly wiring them together by hand and re-testing the resulting circuit, the planar integrated circuit allowed firms to produce circuits directly on the wafer. TI introduced Jack Kilby's hybrid approach in 1959, but it was Fairchild's device, invented by Robert Noyce in 1960 and made possible by Jean Hoerni's development of the planar process in 1959, that became the industry standard.[10] The planar integrated circuit took advantage of the protective and insulating properties of silicon's naturally occurring oxide, silicon dioxide. Entire circuits, rather than single transistors, could be fabricated on the wafer surface, covered with a layer of silicon dioxide and wired together by the deposition of metal directly over the silicon dioxide with contact made by etching through the oxide layer down to the transistor.

4. Moore's Law: The Extendibility of a Dominant Design in Competition and Collaboration

4.1. Moore's Law and the MOSFET Transistor

By the early 1960s, the semiconductor industry had an emerging technological platform, and its market structure began to change as the civilian market for integrated circuits quickly grew to supplant military volumes. Previously, sales to commercial markets had been almost non-existent owing to high costs and manufacturers' unfamiliarity with how to design electronic

[10] While Hoerni and Noyce are credited with the invention of the planar process and monolithic integrated circuit, much of the underlying science – including the discovery of silicon dioxide's protective capabilities – occurred earlier at Bell Labs (Riordan and Hoddeson, 1997 pp 265).

products with ICs. During this time, the military served as an early adopter, mandating the use of integrated circuits in their systems. Now, semiconductor firms began a widespread push to convince commercial electronics users of the maturity and cost effectiveness of integrated circuit technology. This was the larger context in which Gordon Moore, then R&D director at Fairchild, published his fabled paper, "Cramming more components onto integrated circuits," the first formulation of Moore's eponymous "law," in the journal *Electronics* (Moore, 1965). Moore stressed the maturity and stability of silicon integrated circuits, noting that considerable future progress was possible with "only engineering effort" (as opposed to advances achieved through risky or uncertain scientific research). Additionally, Moore projected that the "complexity of minimum component costs" would continue to double annually. His projection was consistent with contemporary assessments by other industry executives.[11] Thanks to an IC price war in the mid-1960s and a concerted marketing campaign by firms like Fairchild, which pioneered the practice of publishing application notes showing engineers how to design electronic products around Fairchild's integrated circuits, sales of ICs to commercial end-users first equaled those to the defense market by 1966. As shown below in Figure 2, thereafter commercial markets would increasingly surpass military demand (Tilton 1971; Lecuyer 2005).

The most radical conjecture in Moore's 1965 paper came from his contention that continued exponential growth in IC component density would open up entirely new applications and technological possibilities at continually decreasing prices. This proclamation signaled a shift in the industry's focus from improving individual device characteristics to increasing circuit density. During the industry's first decade, emphasis had been on improving the operating

[11] An IEEE conference in 1964 featured several industry executives ringing the horn of decreasing prices for integrated circuits. In particular, C. Harry Knowles of Westinghouse, presented complexity-cost charts for integrated circuits that illustrated the same U-curve highlighted by Moore in his 1965 piece. For more, see (Brock & Moore 2006).

characteristics of transistors primarily through new processing techniques and process refinements. At this time, transistors were drop-in replacements for vacuum tubes, offering lower power consumption, reduced size and weight, and improved reliability. However, integrated circuits, argued Moore, offered the possibility of vast new technological capabilities driven by higher complexity of integration. In this paradigm of extendibility, improvements to the now-incorporated transistor would still be pursued but primarily to serve the larger goal of increasing circuit complexity.

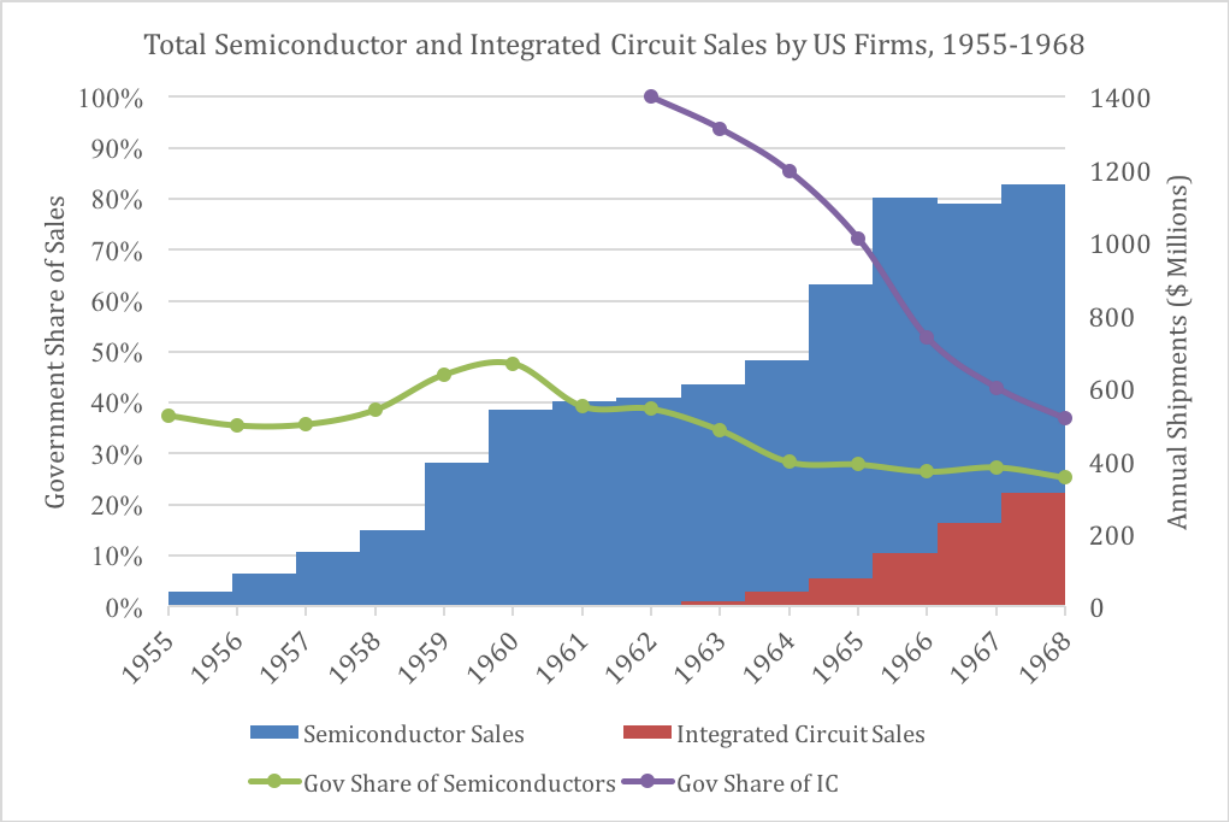

Figure 2 - Total sales of US semiconductor firms and share of sales to Federal Government, 1955-1968. Compiled from data in (Levin 1982).

As a result, a new transistor technology, the **m**etal-**o**xide **s**emiconductor **f**ield-**e**ffect **t**ransistor (MOSFET), better suited to maximizing circuit density than the bipolar junction transistor, emerged thanks to noteworthy research in industry labs, especially at Bell Labs and the RCA's corporate laboratory. Compared to BJTs, early MOSFETs were considerably slower

but offered advantages in miniaturization, manufacturing simplicity, manufacturing cost, and density using a similar set of manufacturing processes employed in making ICs within the BJT paradigm. The MOSFET's historical antecedents go as far back as 1926 to patents describing a field-effect device based on a copper-sulfide material (Edgar 1930). Not until 1960, however, were researchers at Bell Labs able to devise a structure that could reliably overcome surface-effects between the device's different material layers.

The industrial research labs devoted much of the 1960s and early 1970s advancing the art of MOSFET manufacturing (Chih-Tang 1988). MOSFETs first found their way into such products as computer memory and logic chips for calculators.[12] During this early period in MOSFET manufacture, process advantages translated to product advantages and, often, immense commercial gains. Both IC co-inventor Robert Noyce and Gordon Moore figured importantly in this period when they left Fairchild Semiconductor and established a new enterprise, Intel Corporation. In 1969, Intel Corporation became the first firm to use the silicon-gate PMOS process with its launch of the 1101 SRAM (static random-access memory) chip, the company's first successful product. Intel's silicon-gate PMOS process allowed for greater density and lower threshold voltages compared to the aluminum-gate process used by competitors. Similarly, TI spinoff Mostek rose to prominence in the DRAM (dynamic random-access memory) market, briefly obtaining worldwide leadership, due in part to its successful introduction of ion implantation (Bassett 2002).

Following their commercial introduction in the mid-1960s, metal-oxide-semiconductor (MOS) ICS rapidly captured market share through the 1970s. As shown in figure 3 below, MOS

[12] Over 50% of MOS sales in 1973 were estimated to come from two markets, calculators (28% of total MOS sales) and memory (23%). Additionally, the invention of the microprocessor came from Intel's contract with a calculator manufacturer, Busicom.

sales grew from less than 2% of all US firms' IC sales in 1968 to over 52% a decade later in 1978. Through the 1970s, similar to the evolution of the junction transistor in the 1950s with advances in process technologies, the primary technology for MOSFET ICs matured from the relatively slow PMOS to NMOS. By 1980, NMOS-based ICs were the single largest production technology, representing 37% of the overall market for ICs (ICE 1998). At the close of the 1980s, the industry's leading-edge products converged around complementary-MOS (CMOS), a highly energy-efficient form of MOSFET-based circuits, as managing power-density became a primary concern of the market.[13]

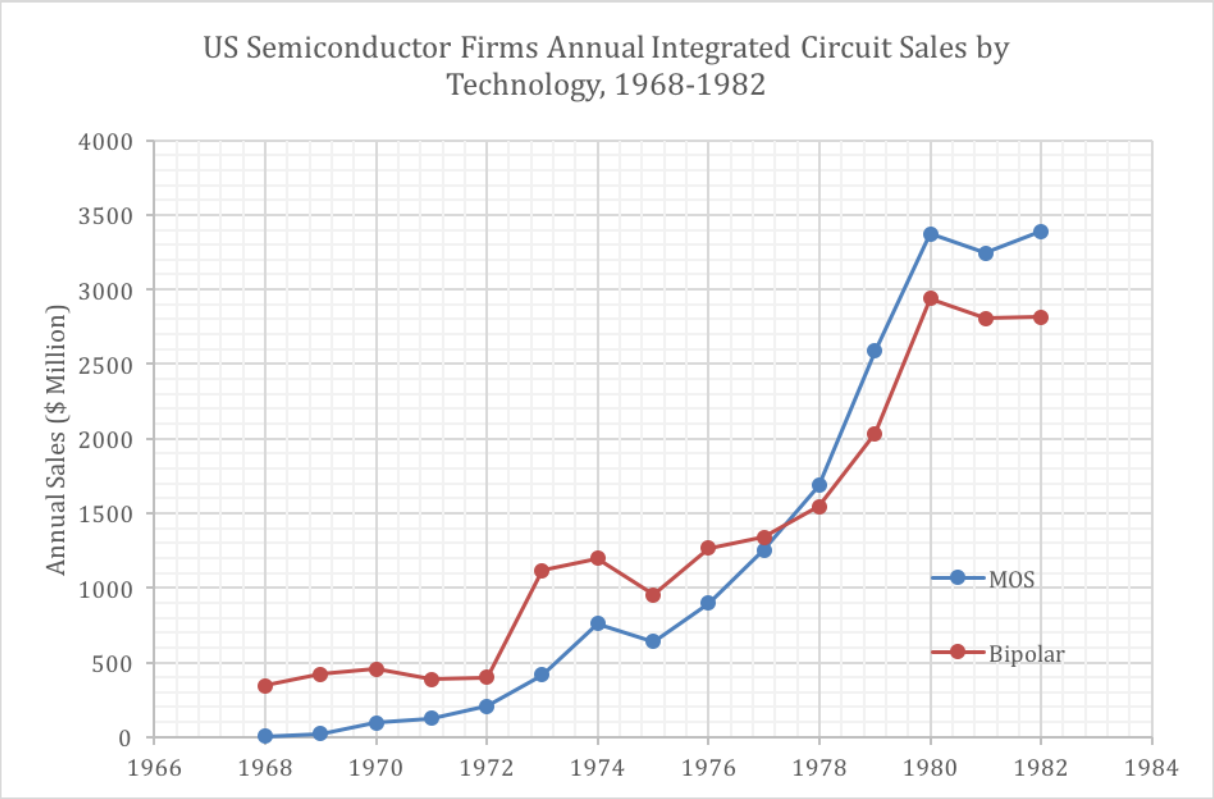

Figure 3 - Annual sales of US semiconductor firms by technology type. Data compiled from Integrated Circuit Engineering Corporation Yearly Status Reports, 1969-1983.

[13] CMOS pairs PMOS transistors with NMOS transistors in a complementary alignment. Only one transistor is "on" at any given time and CMOS transistors only consume power during a switching event (i.e. off to on).

As Gordon Moore had suggested in 1965 (Moore, 1965), increased transistor density was the driver behind achieving higher chip performance at lower cost. Two other publications in the 1970s described the broad technical blueprints the industry followed over the next several decades to maintain that trajectory. "Scaling" – or geometric reduction – of transistors began soon after their commercial introduction and soon raised the specter of new potential limits to device operation —i.e., the appearance of new Chicken Littles in the semiconductor industry barnyard. One early study of limits to MOSFET scaling (Hoeneisen & Mead 1972) highlighted the delicate dance of MOSFET chip designers in balancing various, potential methods of failure as they continued to shrink device features. In response to these serious concerns, researchers at IBM developed a general theory for how to shrink MOSFET transistors while avoiding so-called "short-channel" effects (Dennard et al. 1974).[14] Dennard Scaling, as it came to be known after IBM Research's theoretician Robert H. Dennard, summarized the parameters available for toggling (dimension, voltage, and doping) while identifying challenges that arose with continued scaling (minimum gate oxide thickness, interconnect resistance, and non-scaling of the subthreshold slope). Through the early 2000s Dennard Scaling provided the primary pathway for progress within the MOSFET integrated circuit paradigm: the toggling parameters would be the primary levers for enabling transistor scaling and thus progress in semiconductors. Simultaneously, the feared challenges were overcome by engineering around the complications posed by the latter three scaling parameters.[15]

[14] The findings in Dennard et al (1974) were not new theory. The relationship between device (transistor) size and operational characteristics was present even in Shockley's initial theory of the p-n junction transistor (Shockley, 1951). Furthermore, Moore (1965) specifically mentioned the decrease in power per unit area with reductions in feature size. "Dennard Scaling" codified this knowledge and experimentally proved the extendibility of MOSFET transistors.

[15] In practice the industry has used a more general form of scaling to maximize performance by scaling voltage slower than doping and dimension. This allowed for devices to operate at faster speeds and, in the earlier years, aided with system compatibility by keeping transistors at 5V (Baccarani et al. 1984),

Transistor scaling was only one aspect of increased transistor density, however. In 1975, at a conference, Moore revisited the trend toward higher integrated circuit complexity. First he dissected the trend toward higher chip densities into three components: larger die (chip) sizes, "circuit cleverness," and transistor shrinkage (Moore 1975). Moore concluded that, while die size increases and transistor shrinkage would continue unabated, he expected density gains from circuit cleverness to tail off in the coming years as new devices with optimal layouts – charge-coupled devices – hit the market.[16] Consequently, he observed, the doubling of densities would slow from occurring annually to taking place every two years. As to ICs reaching any limits in overall device scaling, Moore was coy. He noted only that "we are still far from the minimum device sizes [being] limited by such fundamental considerations as the charge on the electron or the atomic structure of matter."

4.2. Chicken Littles in the 1960s and 1970s

Moore's confidence on the limits to extending the silicon integrated-circuit paradigm still being far off was not an industry-wide consensus. Writing in 1965, the same year as Moore's *Electronics* publication, Jack Morton, then Bell Labs' Vice President of Electronic Technology, argued that while the integrated-circuit offered a short-term solution to the problem of integration, in the long-term the industry needed to look to basic scientific advances to create circuits built on "functional devices" (Morton 1965). Morton envisioned a parallel trajectory for semiconductors driven not by increasing complexity of integrated circuits enabled by geometric scaling but instead by the simplification of circuits through functional devices made possible by fundamental advances in physics.

[16] In fact, Moore predicted this slowdown to occur within a decade. The change in slope took place almost concurrently with his publication. Charge-coupled devices never reached market significance.

This alternate trajectory was kept alive through the 1970s and 1980s, as a steady chorus of Chicken Littles from the industry's basic research labs raised alarms about the limits to integrated-circuit extendibility in the current paradigm. Surveys of potential limits published by IBM researchers hinted at the organization's interest in alternatives to silicon integrated circuits. R.W. Keyes argued that "based on extrapolation of present technology" limits to speed were only an order of magnitude away and thus "[p]rogress beyond this point can only be made by radical deviations from the current lines of development" (Keyes 1969). Based on promising experiments from the early 1960s, in 1967 IBM launched a computing project based on superconducting Josephson-Junctions. IBM researchers argued the technology offered the potential for faster circuits with considerably lower power dissipation relative to silicon integrated-circuits. However, after fifteen years of extensive R&D the project was ignominiously cancelled in 1983. Technical difficulties in delivering a Josephson-Junction computer remained, and the continued rapid advance in silicon integrated-circuits had all but eroded the potential for the alternate technology's performance advantages to gain market traction.[17]

Military funding and research labs also helped support alternative and far-looking technologies during this period. The Office of Naval Research first established the Ultra Submicron Electronics Research (USER) program in the 1970s. That program's stated goal was developing "an electronics technology based on devices with 20 angstrom feature sizes," well below the perceived limits for silicon integrated circuits (Cooper 2007). Groups at ONR and

[17] For a review of the program see (Anacker 1980). For an internal history of IBM's work in superconductors see Gallagher, Harris, and Ketchen (2012).

NRL would also support newly awakened visions for "molecular electronics" that did not find strong support inside the industry's research labs.[18]

At Texas Instruments, meanwhile, the move by George Heilmeier from DARPA sparked new lines of inquiry into alternatives to silicon integrated circuits. Heilmeier moved to TI in 1977 and named the firm's Chief Technology Officer in 1983 and starting in the mid-1980s, TI began a program investigating "quantum effects" devices. The program, which was at least partly funded by the Department of Defense, was positioned as a response to predictions of fundamental limitations facing silicon integrated-circuits at or around 0.1 micron (100 nm) device lengths. By this period many felt that newer lithographic technologies would offer a path beyond optical lithography's limits. Similar to IBM's 1967 effort, however, the program's researchers argued that continued progress in microelectronics in the existing paradigm would be inhibited by degradation of device operation at smaller geometries due to quantum effects and limits to density imposed by device interconnects. TI researchers proposed both new device structures, resonant tunneling transistors, and architectures, cellular automata, all outside the existing paradigm and designed to circumvent the device and interconnect limits envisioned.[19] By the early 1990s, however, TI had closed its Central Research Lab and the program's researchers had found new homes in academia. In this new university environment many of the investigators and their concepts continued to receive funding through DAPRA's ULTRA Program.

---

[18] Hungsub Choi and Cyrus C. M. Mody, "The Long History of Molecular Electronics: Microelectronics Origins of Nanotechnology, "Social *Studies of Science* 39, 1, (2009) 11-50.

[19] See (Bate 1986) and (Bate et al. 1987) for a description of the motivation behind the program. Bate (1986) specifically references work from TI's device group foreseeing limits to MOS around 0.5 micron (Chatterjee et al. 1983).

4.3. Three Decades of Dennard Scaling

Despite the considerable research and investment into silicon integrated circuit alternatives, the 1970s marked the first of three decades of Dennard-Scaling-driven rapid progress in semiconductors. As the industry continued within this paradigm for technology advancement, with its focus on process improvement, international competition, primarily from Japan, and vertical disintegration reshaped the business of semiconductors. Unlike the early 1960s, when uptake of leading-edge integrated circuits was governed by systems-integrators (mostly military vendors), the industry's technological pacesetter in the 1970s was a commodity product, the DRAM chip (Dynamic Random Access Memory DRAM). DRAM chips remain integral to Von Neumann based computing architectures to this day – programs and data are loaded into memory and accessed as needed for computation. The expansion of the commercial computing market during the 1970s was governed by the availability of the next generation of DRAM chips, which arrived in roughly three-year intervals. For much of the 1970s and 1980s, DRAMs were the industry's single largest product. Their relatively simple design – a vast, repeating array of the 1-transistor/capacitor design invented by Dennard in 1968 – meant the product was the perfect test bed for new process technologies. The importance of DRAM performance to the commercial computing market also meant that firms first to market with the next-generation DRAM, usually enabled by a step-down in lithographic feature size, would capture significant monopoly rents.

That competition in DRAMs focused around process innovation helped enable its capture by Japanese firms. Following two decades of active industrial policy[20] in the computing industry,

[20] In 1957, Japan's Ministry of International Trade and Industry (MITI) became the shepherd of the country's electronics industry. In the early 1970s, MITI would re-organize Japan's computing industry which had been wiped out by the introduction of the wildly successful IBM 370 (Flamm, 1988).

Japan's Ministry of International Trade and Industry (MITI) launched the VLSI Program: Japan's VLSI program was a collaborative R&D project centered around joint laboratories totaling $288 million in funding from 1976 through 1980. The program's goal was to develop leading-edge semiconductor technology to provide Japan's computing industry a competitive advantage (Ouchi 1984). In retrospect the VLSI program was not a successful horizontal collaboration as its member firms remained mutually suspicious and at times intransigent (Sakakibara 1983). Rather, the program's primary benefit was the improvement of Japanese semiconductor equipment supplier capabilities. Prior to the VLSI program, Japanese manufacturers relied heavily on US vendors for semiconductor equipment; one estimate suggests that a quarter to a third of all VLSI funds was spent purchasing US equipment (Sakakibara 1983). The success of the program in developing sub- micron technology bolstered the technical capabilities of Japanese equipment suppliers, especially in key technologies like lithography, and decreased dependence on foreign equipment suppliers from 75 percent in 1977 to 50 percent in 1980 (Sakakibara 1983). In addition to these advances in Japanese supplier production equipment capabilities, the diversified and vertically integrated nature of Japanese semiconductor manufacturers enabled larger investments in production facilities combined with their use of more conservative designs helped Japanese firms capture over 80% of the DRAM market by the mid-1980s. In large part through their dominance of the DRAM market, Japanese firms gained the global lead in semiconductors. All but one U.S. DRAM manufacturer left the DRAM market and focused elsewhere or exited the industry.

4.3.1. New Institutions to Address International Competition

American semiconductor manufacturers began organizing against Japanese competition in 1977 with the formation of the Semiconductor Industry Association. Five firms established the

SIA to "represent the common interests of the industry in matters of trade and governmental policy, and to bring emphasis on common problems and opportunities such as safety and trade statistics" (Semiconductor Industry Association 1981)[21]. The nascent SIA was an offshoot of the umbrella trade group, the Electronics Industry Association, and early initiatives put forth by the SIA targeting Japanese trade and dumping tactics failed in part because of opposition from large chip purchasers – e.g., computer manufacturers – who preferred low prices. These early failures required the SIA to expand its membership to include vertically integrated captive producers, firms like IBM, Hewlett-Packard, AT&T, and Digital Equipment Corporation (Browning & Shetler 2000). Managing the at-times divergent interests of its newly expanded membership would shape the SIA's agenda. Inclusion of vertically integrated captive producers in the SIA gave the organization more political clout but also tempered the group's policy proposals to be more acceptable to its largest members (Yoffie 1988).

[21] The five firms: AMD, Fairchild, Intel, Motorola, and National Semiconductor.

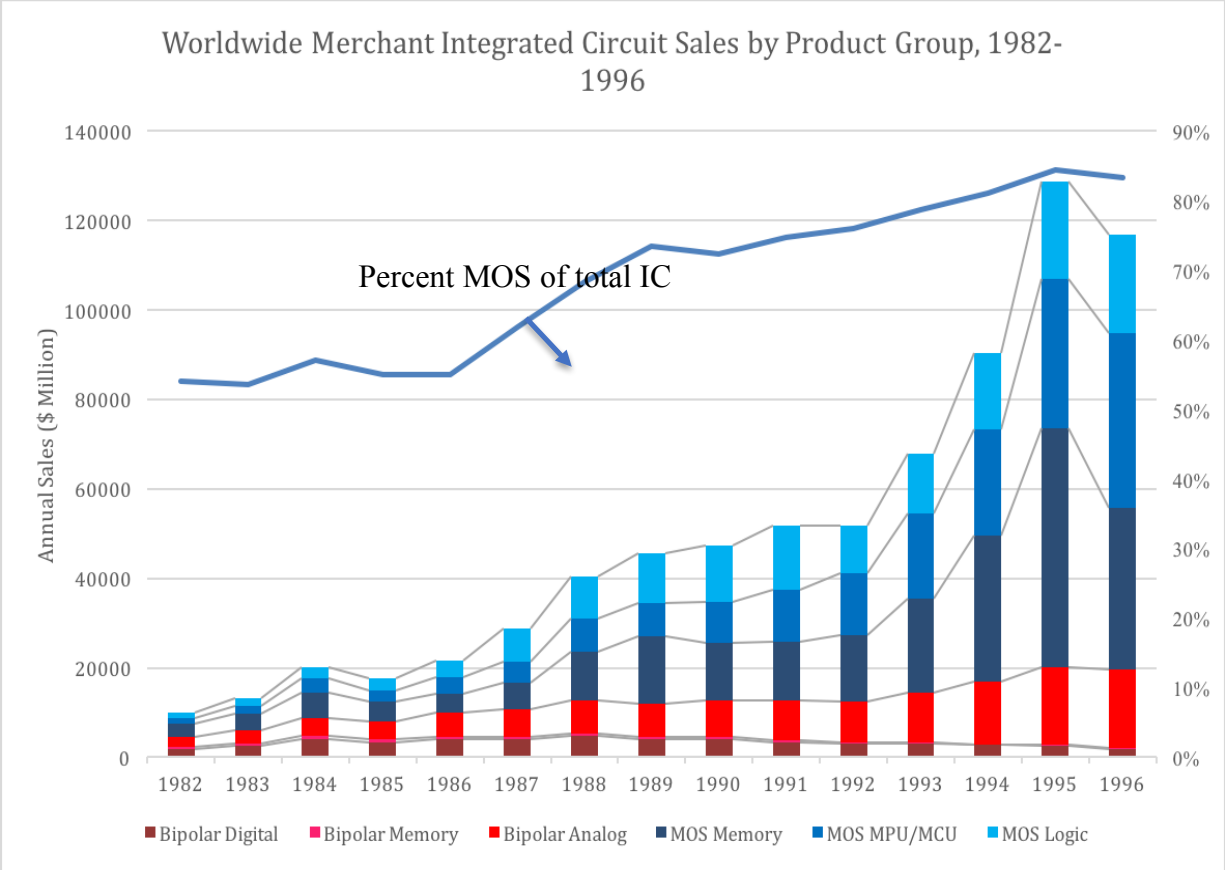

Figure 4 - Breakdown of worldwide merchant integrated circuit sales by product group, 1982-1996. Data compiled from Integrated Circuit Engineering Corporation annual "Status Reports", 1982-1997.

When Japanese firms began leapfrogging American competitors with the introduction of the 64K DRAM, the American computer and semiconductor industry began embracing the idea of collaborative research, despite long-standing anti-trust statutes that limited the types of cooperation available to US firms. Two collaborative research ventures, both funded entirely by member firms, were launched in 1982, the Semiconductor Research Corporation (SRC) in January and the Microelectronics and Computer Technology Corporation (MCC) in August. While the MCC was initially launched with a much larger budget and considerable fanfare, it would slowly fizzle out and be viewed as a failure.[22] In establishing the SRC, the SIA explicitly

[22] For a full history of the MCC see Gibson, D.V. and Rogers, E.M., 1994. *R & D collaboration on trial: The Microelectronics and Computer Technology Corporation*. Harvard Business Press.

mentioned Japanese and European industrial policy and the decoupling of the industry's technology trajectory from military needs as motivations behind the program (SRC 1983). Emblematic of the growing disconnect between military needs and commercial integrated circuit technology was the Department of Defense's VHISIC program. Launched in 1979, VHISIC focused primarily on the development of application specific integrated circuits for defense applications, which did not coincide with the advancement needs of the more general, commercial semiconductor technology.[23]

The SRC focused on elevating the industry's needs to areas deemed worthy of study by leading academic researchers. Although US firms had identified manufacturing as a critical area for improvement, there was little university research in the area in 1982. Robert Burger, the SRC's longtime Chief Scientist noted that the SRC included "sciences" in the title of its "Manufacturing Sciences" program to "emphasize its 'intellectual content,'" and to increase appeal to academics (Burger 2000). The focus on manufacturing research in the initial SRC agenda reflected the industry's prevailing challenge and was also popular amongst the consortium's member companies. Japanese firms had taken control of the semiconductor market in the 1980s through their advantages in manufacturing efficiency and quality, achieving higher line yields on new DRAM products ahead of American competitors. SRC's 1983 annual report reads, "SRC member companies almost unanimously agree that there is no more important task for the SRC than elevating semiconductor manufacturing technology from an art to a science." Two proposals with this focus, CERES in 1983 and Project Leapfrog in 1984, were, however, rejected by the SIA board. Despite these rejections, the industry's focus remained on

[23] After the launch of VHISIC, although military R&D funding for integrated circuits rose, the proportion of military R&D funding in silicon integrated circuits decreased. DARPA in particular focused on efforts to commercialize GaAs technology (ICE 1985).

manufacturing competitiveness, and a series of Manufacturing Competitiveness Panels during 1985 and 1986 laid the groundwork for what would become SEMATECH (Burger, 2000).[24]

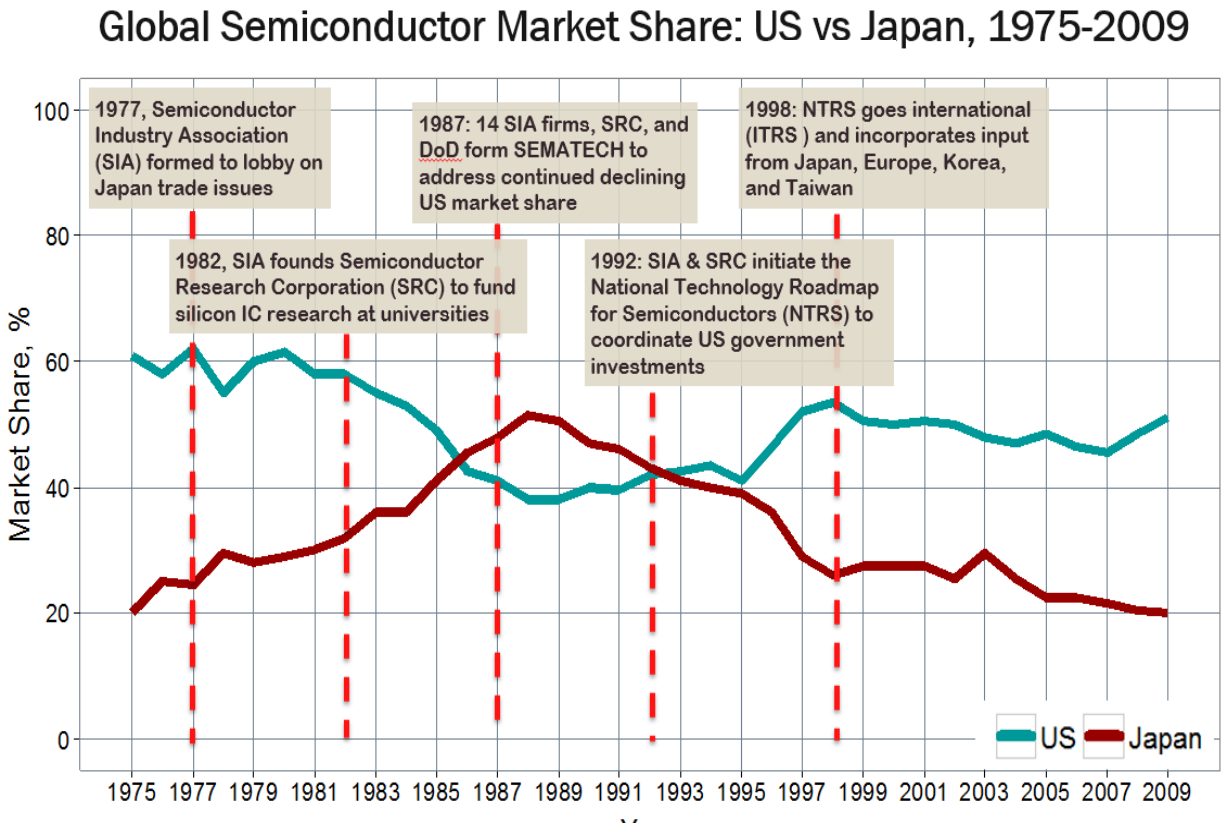

Figure 5 - Global market share of US and Japanese semiconductor firms. Data source: SIA

Despite these early efforts, US semiconductor manufacturers continued to cede market share to Japanese firms. Concern over the future of the US semiconductor industry, and in particular semiconductor equipment suppliers of key technologies such as lithography, reached a fever pitch in 1987. During 1986, Sandy Kane, an IBM vice president, toured the country delivering an "obituary" for the US semiconductor industry. George Scalise, then head of the SIA, committed $100,000 from the SRC at an SIA board meeting to study the issues with US

[24] Burger gives a more complete chronology of these proposals and their run-up to SEMATECH.

semiconductor manufacturing, and Charles Sporck, CEO of National Semiconductor Corporation, spent the year meeting with industry executives, academics, and politicians to devise a strategy for a collaborative venture (Browning & Shetler 2000, p.17). Meanwhile, the government, led by the Defense Department's Defense Science Board (DSB), issued two separate reports about the importance of microelectronics and semiconductor manufacturing for defense readiness (Browning & Shetler 2000, p.19).

In response to commercial and military concerns, 14 US semiconductor manufacturers formed SEMATECH in 1987 and Congress authorized $100 million in annual funding for 5 years through DARPA. Similar to MCC and Japan's VLSI, SEMATECH's participants initially failed to agree on the consortium's strategy. Member firms had different visions for SEMATECH depending on their relative standing in the industry's pecking order. Smaller firms wanted a process development focus, larger firms preferred a focus on improving equipment supplier capabilities, and firms with large military contracts mirrored DARPA's preference for flexible manufacturing of ASICs (Browning & Shetler 2000).

SEMATECH eventually forewent horizontal collaboration and refocused the consortium on improving US equipment supplier capabilities. For manufacturers, the threat of Japanese domination of key equipment categories, like lithography, made this approach a necessity. Japanese equipment suppliers were suspected of withholding new technologies from US manufacturers for a generation, giving Japanese manufacturers a sizable advantage (Browning & Shetler 2000, pp.105, 125). Meanwhile, US equipment suppliers, largely shut out of foreign markets, needed to remain competitive to maintain market share within the US (Carayannis & Alexander 2004). Data indicate that US equipment suppliers were successful in improving their competitiveness during this period. Equipment improvements cut the Japanese yield advantage

from 50% in 1985 to 9% in 1991 (Spencer & Grindley 1993). By 1991, US semiconductor firms purchased 70% of their equipment from US suppliers helping US equipment manufacturers improve their market share from 45% in 1990 to 51% in 1992 (Spencer & Grindley 1993).

By 1993 U.S. semiconductor firms had once again surpassed Japanese manufacturers in global market share. A Japanese recession and the emergence of low-cost competition in Korea and Taiwan helped erode Japanese market share (Carayannis & Alexander 2004). While US companies were more competitive, their recovery is only partially explainable by SEMATECH. US firms benefited from a combination of their decisions to exit the DRAM market and focus on higher margin products, changes in US-Japanese trade policy, and improvements in manufacturing capabilities (Grindley et al. 1994; Macher et al. 1998; Carayannis & Alexander 2004). The reversal of fortunes for US semiconductor firms was also closely linked with the upheaval in the market for computing. Personal computers, first introduced in the late 1970s, captured increasing market share through the 1980s and surpassed traditional mainframe computers by the early 1990s. Whereas mainframe computing's leading firms were mostly vertically integrated, often utilizing proprietary chips and software for their own product offerings, the personal computing market relied primarily on commodity components that helped ensure interoperability of software. As a result, semiconductor and software firms working in collaboration, such as the "WINTEL" collaboration between Intel and Microsoft, captured a significant portion of the profits in the vertically disintegrated personal computer market.[25]

During this same decade when the U.S. lost and regained it competitive advantage against Japan, the US semiconductor industry underwent its first wave of vertical disintegration. The business of semiconductor manufacturing split into product design and process

[25] The WINTEL duopoly saw its combined revenues rise from $5.1 billion in 1990 to $56.7 billion by 2000.

development. This split was made possible by the stabilization of MOSFET manufacturing technologies and improvements in electronic design automation (EDA) software that allowed designers to incorporate knowledge of manufacturing capabilities into circuit designs (Macher et al. 1999; Macher & Mowery 2004). "Fabless" semiconductor firms—companies without their own manufacturing facilities—entered the market in the 1980s, initially utilizing spare capacity at integrated device manufacturers.[26] Between 1978 to 1987, two-thirds of U.S. semiconductor startups founded owned no manufacturing facilities (Angel 1990).

At the same time, the IBM-driven open architecture of personal computers quickly converged to the dominant "Wintel" standard, resulting in the largest market for semiconductors being standardized memory (DRAM) and logic (microprocessors) chips. Chips in both of these markets were dominated by MOS technologies and their rapid growth drove the emergence of CMOS as the industry's dominant technology. In contrast to the semiconductor start-ups, the large firms competing in these two markets tended not to vertically disintegrate because complementarities between design and manufacturing provided keys to their competitive advantage (Macher, 2004). These integrated device manufacturers were also more likely to patent "systemic innovations" than their vertically disintegrated peers (Kapoor 2013). Firms competing in these commodity chips drove process innovations that spilled over to the rest of the industry through the industry's collaborative institutions (Lim 2009). This dynamic propelled microprocessors to the process frontier for semiconductors. The centrality of the microprocessor to the capabilities of personal computers ignited intense competition to deliver the fastest

[26] Captive producers also began to outsource their production to foundries during this time. For example, Intel and Burroughs Corporation struck such a deal in 1982 (ICE 1984)

microprocessors, leading to an acceleration of Moore's Law relative to the industry's roadmap.[27] Technically this acceleration was accomplished by scaling voltage more conservatively than predicted by Dennard Scaling, allowing for faster speeds but increasing a chip's overall power consumption. Process advantages also helped leading microprocessor firms fend off competition from alternative architectures (Hennessy & Jouppi 1991).

4.3.2. Vertical Disintegration and Collaborative R&D

Finally, along with vertical disintegration and market shift in the 1980s, increasing development costs, shortening product cycles, and more sophisticated competition both domestically and, increasingly, from Japan limited the ability of firms to capture the social and private returns from their R&D investments (Mowery & Rosenberg 1991). Most of the new fabless semiconductor firms shied away from investing in more traditional forms of industrial research, including "in-house" basic or fundamental research. Although overall R&D spending by the U.S. semiconductor industry increased dramatically in the 1980s, the central research labs of the large semiconductor manufacturers like IBM, TI, RCA, and AT&T faced major headwinds as the decade progressed. Most firms enacted significant cuts to their basic research budgets, refocusing of projects away from "blue-sky" research toward applied research in support of their existing businesses (Macher et al. 1998). Research at Bell Labs shifted after 1983 when AT&T was broken up. The acquisition of RCA by General Electric in 1986 —solely for obtaining control of the National Broadcasting Company (NBC)—led immediately to GE's abandonment of RCA's illustrious Sarnoff Laboratory in Princeton, New Jersey. IBM's struggles in the early

[27] It is worth nothing that many of the additional transistors on a microprocessor were devoted to on-chip memory, not logic functions.

1990s brought a complete overhaul of its research strategy: although basic research remained at IBM, it was more directed and tied to commercial products. In the mid-1990s, Texas Instruments completed shuttered its central research organization. Meanwhile, two of the industry's emerging juggernauts, Motorola and Intel, eschewed central research organizations entirely.[28] As the industry's leading firms scaled back their basic research budgets or failed to entertain holding research organizations at all, the military, principally through its Defense Advanced Research Projects Agency (DARPA), played the primary funding and coordinating role in non-CMOS semiconductors and alternatives to the tried-and-true optical lithography used to manufacture CMOS chips.[29]

In 1988, the year following SIA's founding of SEMATECH, Congress authorized the National Advisory Committee on Semiconductors (NACS) to "recommend appropriate actions that support the national semiconductor strategy" (National Advisory Committee on Semiconductor Research and Development Act of 1988)[30]. The NACS charter explicitly recognized the importance of semiconductor technology to the wider electronics industry and military readiness as justifications for increased government-industry cooperation in restoring competitiveness. The NACS issued a series of reports in the early 1990s laying out a comprehensive industrial policy to improve the semiconductor industry's competitiveness (NACS 1991b; NACS 1992a; NACS 1992b). The committee's wide-ranging recommendations – including tax and education reforms – explicitly called for the government to support the

[28] From its founding, Intel went without a central research laboratory and instead focused on experimentation on the factory floor. Under Robert Noyce's principle of least (or minimum) information. See Moore in Rosenbloom & Spencer, 1995.

[29] The military had a long-running interest in "leapfrogging" conventional silicon technology. For a history, see "The Long History of Molecular Electronics." During the 1990s, DARPA's ULTRA Electronics and "Advanced Lithography" programs were major funding sources for alternatives to CMOS and next generation lithography approaches.

[30] https://www.law.cornell.edu/uscode/text/15/4632

development and transfer of semiconductor process technologies through increased SEMATECH funding and renewed efforts to commercialize X-ray lithography technology from the Department of Energy's National Labs. Many in the industry believed optical lithography was approaching fundamental limits and Japanese firms, which already held the lead in conventional lithography equipment, were also perceived to be ahead in the development of emerging X-ray lithography techniques. NACS recommendations also shaped government investment in semiconductor R&D; DARPA's "Advanced Lithography Program" funded both DUV and X-ray lithography approaches through the 1990s in close coordination with industry.[31]

Meanwhile, the predictability of semiconductor advancement – still following the Moore (1975) and Dennard et al. (1974) blueprints – enabled long-ranging forecasts extrapolated from the industry's existing trajectory. The first collaborative roadmaps in the US grew out of collaborative research aimed at leapfrogging the newly ascendant Japanese competition[32]: SRC's "1994 goals" – issued in 1984 – called for 0.25 micron linewidths and 256 mb DRAMs.[33] These goals were further codified in 1991, with "Micro Tech 2000", the US semiconductor industry's first widely-publicized collaborative technical roadmap. The White House OSTP, under the aegis of the NACS, had organized a conference of industry, government, and academic researchers "to identify the most critical efforts that should be undertaken to develop the [0.12 micron] manufacturing process" by the year 2000 (NACS 1991a). MicroTech 2000 focused on

[31] DARPA's "Advanced Lithography Program" continued through the 1990s and into the early 2000s. For DARPA work on next generation lithography see "Darpa ends litho aid" (2005)

[32] This predictable extendibility was also the basis of Japan's VLSI program, which helped Japanese firms capture the DRAM market in the 1980s.

[33] The SRC issued updated projections for 2001 in 1989, building on the 10-year goals from 1984. "The technology push for 1 gigabit DRAM toward the turn of the century predicts the minimum-feature size of 0.15um and will require sophisticated lithography such as X-rays. For microprocessors, the gate delays are projected to be 15 ps with a 200MHz clock frequency. Wafer size will exceed 250 mm. Single-wafer and fab cost is expected to exceed $1 thousand and $1 billion, respectively." SRC Newsletter, November 1989 (From Schaller 2004)

challenges to overcome in key process technologies to achieve these goals. According to Schaller (2004), the MicroTech 2000 roadmap's audacious goals struck many in the industry as unachievable and too expensive. Around the time of its publication, industry leaders actually called for an "Apollo-like" program to achieve the acceleration projected in MicroTech 2000.[34] No such project transpired.

Instead, beginning in 1992, the roadmap came under the purview of the SIA as the National Technology Roadmap for Semiconductors (NTRS). In taking ownership of the roadmap, the SIA expanded beyond simply a trade organization to a technology coordinating organization. Organization of the roadmap fell to the group's newly created "Technology Committee" chaired initially by Gordon Moore. The nature of the NTRS differed little from MicroTech 2000. The roadmap continued to be based on projections of technical needs for leading-edge commodity products, primarily microprocessors and DRAM. The NTRS moved to a 15-year horizon, encompassing five separate three-year technology generations. This cadence essentially mimicked the DRAM "4x/3" trajectory established in the 1980s. One difference was the explicit reference to CMOS as the industry's standard technology.[35] In fact, the early roadmaps nearly perfectly continued to extrapolate the existing Moore/Dennard trajectory through the end of the 21st century's first decade. See figure 6 below for a comparison of these

[34] See Shaller (2004) page 486. The news articles he cites are: Andrew Pollack, "Apollo-Type Program Is Envisioned for Chips," New York Times, Feb 21, 1991, Business Section, 2, and George F. Watson, Alfred Rosenblat, "Advisory Committee Report: 'Apollo'-type program for chip industry proposed," The Institute: Inside IEEE, Vol. 15, No. 4, April 1991, 1, 8.

[35] This coincided with CMOS taking nearly complete market share of leading logic and memory products. IBM introduced its first all-CMOS computer in 1996.

projections with actual CPU data. Although gate length reductions actually occurred faster than every projection, die sie and CPU frequency leveled off much earlier than anticipated.

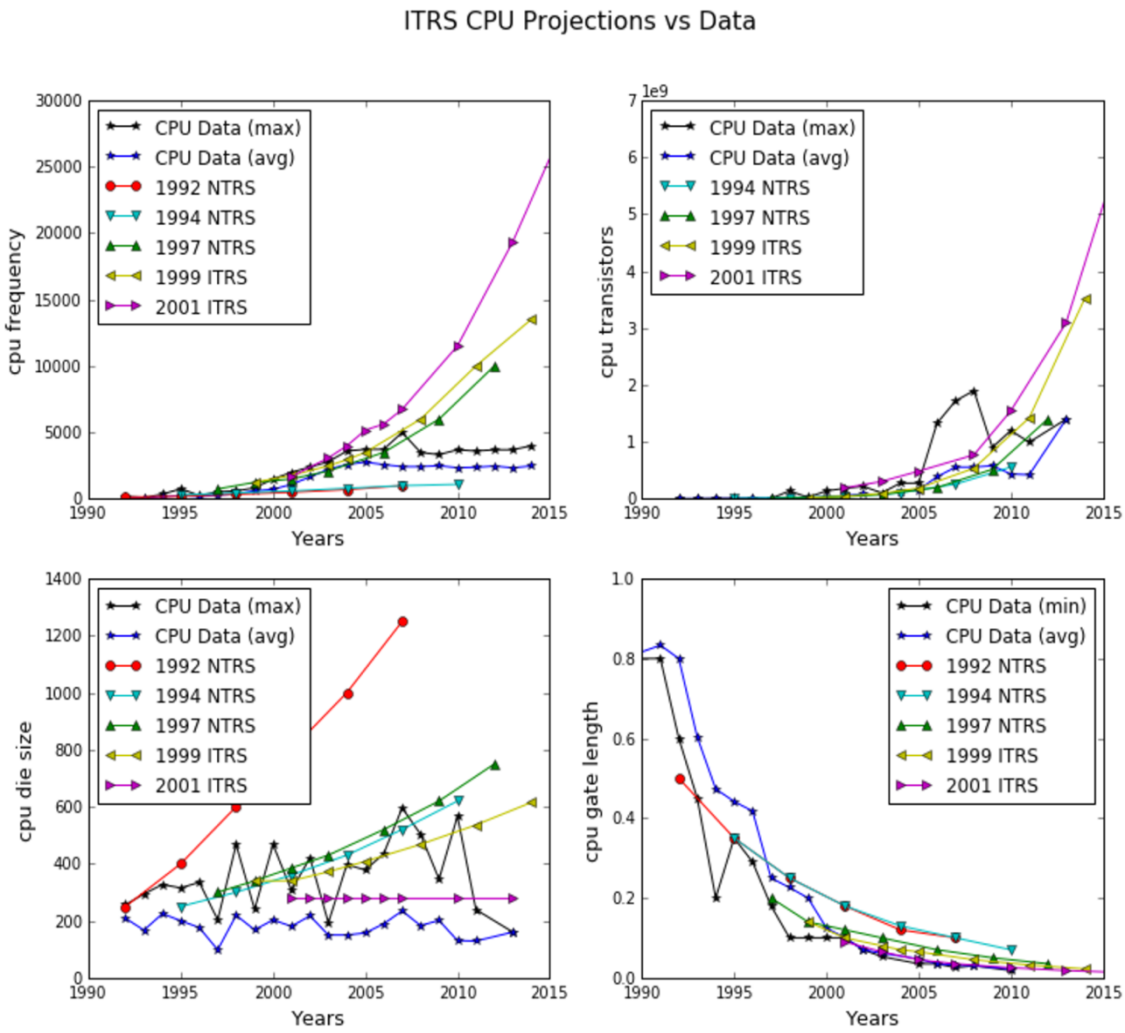

Figure 6 - Comparisons of ITRS projections for CPUs and actual CPU data. Source: NTRS, 1992; NTRS, 1994; NTRS, 1997; ITRS, 1999; ITRS, 2001; Danowitz et al. 2012.

4.3.3. Roadmaps and Codifying Moore's Law

Despite programs like TI's suggesting the contrary, the early roadmaps were unequivocal on the question of limits to scaling. MicroTech 2000's preface stated in 1991, "no fundamental technical obstacles are likely to prevent the continued rapid pace of semiconductor technology advancement. Semiconductor technology will continue to be pushed ahead by a combination of

evolutionary, incremental advances in technology, and revolutionary innovations." The issue of limits and the need for novel devices returned, however, by the 1994 roadmap. Voltage scaling was introduced in the early 1990s as a way to engineer around limits to the electric-field inside the transistor. The runway to continue reducing voltage, however, was limited by the device's subthreshold slope. The roadmap's authors recognized this dilemma and encouraged research into "novel devices" in a section titled "Needs Not Addressed." This section ends with the following warning, "[t]here is also concern that the semiconductor community is not doing enough to understand practical device scaling limits (related to atomic-scale variations, quantum effects, tunneling, shallow junctions, etc.)."[36] In the 1997 roadmap, however, the topic is barely addressed, with only brief mention is made of the need for research into "post-shrink" devices.

Over the course of the 1990s the industry-led roadmap came to define the direction and focus of research efforts for the industry's suppliers and research organizations, including SRC and SEMATECH (Gargini 2000). The SRC re-organized its research structure to mimic the 1994 roadmap's chapters. Academic researchers also began to rely on the roadmap, frequently citing issues raised in the roadmap in research grant solicitations. As such, the roadmap served to pull the focus of academic research toward the industry's most pressing, short-term problems (Burger 2000; Spencer & Seidel 2004). At the same time, with the decline of the industry's central research labs it was unclear if the industry could fund much-needed longer-term research. An analysis commissioned by SIA and SRC following the release of the 1994 NTRS found industry and government funding for long-term research insufficient to address the technology obstacles

[36] These are, essentially, the same limits raised as far back as 1962 by Wallmark and Marcus and reiterated by Hoenesian and Mead in 1972.

identified by the NTRS. The analysts recommended that universities should be used to address a considerable portion of the projected "research funding gap" (Bodway et al. 1995).

In late 1994, the Semiconductor Technology Council was established with representatives from government (NSF and Departments of Defense, Energy and Commerce), industry (semiconductor manufacturers, suppliers and users), and academia "to help guide research and development in semiconductors" as well as "link assessment by the semiconductor industry and national security needs for cooperative investments … and align industry and government contributions for new semiconductor research and development efforts" (STC 1996). With the end of SEMATECH's federal funding in 1996, the council proposed a new program for funding university research. Semiconductor manufacturers, equipment suppliers, and the federal government (initially through DUSD[37] and then DARPA) jointly funded a new SRC subsidiary, the Microelectronics Advanced Research Corporation (MARCO) to oversee the Focus Center Research Program (FCRP). The aim of FCRP was to fund university research centers to "seek creative options for the solution of key technology challenges so that the industry can keep pace with the cadence of Moore's Law" (SRC 1998). FCRP, unlike the SRC's existing research program, was to be driven by the research interests of professors who had long-standing research relationships with the industry. Given the existing inertia and incentives, however, for much of the following decade, MARCO's research funding remained within the Si-CMOS paradigm.

The resurgence of U.S. semiconductor manufacturers in the early 1990s engendered changes in the membership and focus of the collaborative institutions the industry had founded in the 1980s. The opening up of these institutions to foreign membership signaled a shift in focus

[37] US Department of Defense's Undersecretary of Defense

away from international competition toward maintaining technology leadership as U.S. firms began to collaborate with the international community on defining research needs and technology directions. In 1997, American and Japanese firms coordinated through the International 300 MM Initiative to define standards for the shift to 300 mm wafers. This effort was a notable departure from the nationally-focused support for semiconductor equipment firms by both nations in the 1970s and 1980s. In 1998, the industry's roadmap was jointly sponsored by the semiconductor associations of the USA, Japan, Korea, and Europe and renamed the International Technology Roadmap for Semiconductors (ITRS).[38] By 1999, both SEMATECH and SRC had admitted international firms as members. Into the early 2000s, industry leaders mentioned the possibility of "super-consortia" collaborating across international borders (National Research Council 2003).

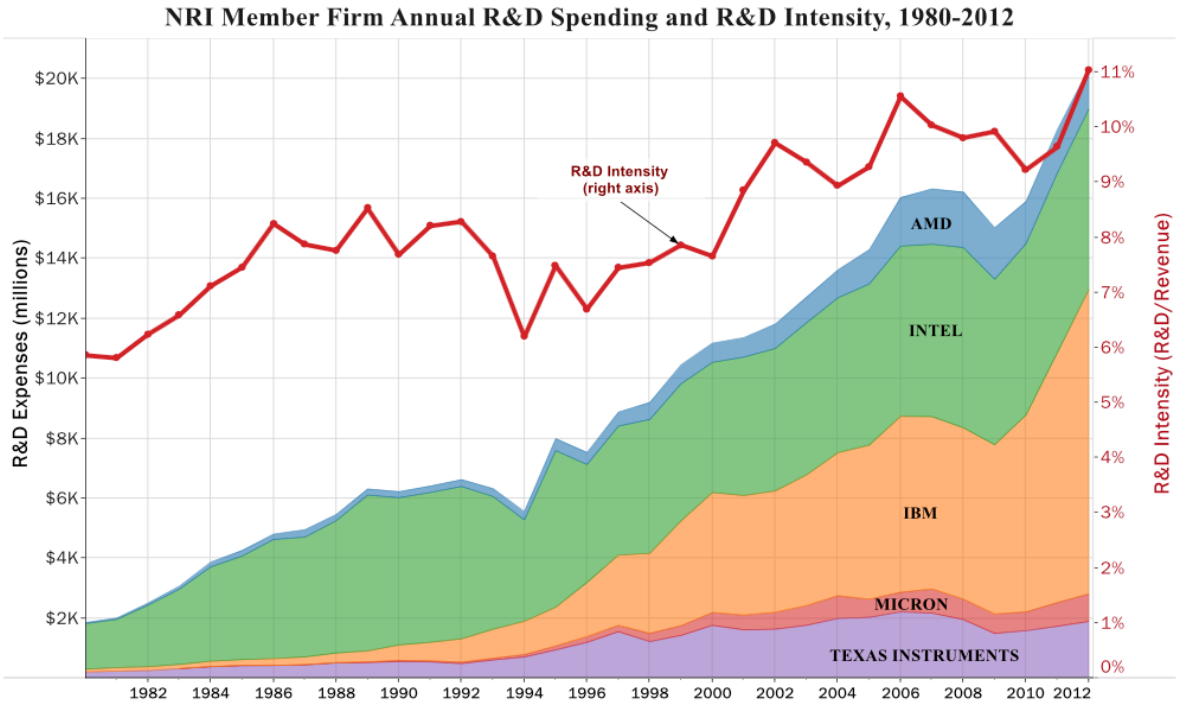

Figure 7 - NRI member firm R&D spending began accelerating in the late 1990s as semiconductor manufacturers had to grapple with new technical challenges. Source: (Standard & Poor's 2013)

[38] According to Paolo Gargini, a key architect of the NTRS and the head of the USA's delegation to the ITRS, the first international collaboration on the roadmap required a revision to over 50% of the previous year's metrics (Gargini, 2015).

5. Beyond-Moore: Divergent Paths and a Return to Experimentation

5.1. Materials Challenges and Equivalent Scaling

The turn of the century, however, brought increasingly difficult technology challenges for the entire industry. By the late 1990s and early 2000s, the industry began to reckon with many of the limits first highlighted several decades earlier. In a widely cited *Science* article an Intel researcher highlighted three challenges described in the 1998 ITRS as "roadblocks" with "no known solutions": dopant clustering, electron tunneling through the gate oxide, and dopant distribution (Packan 1999).[39] Yet semiconductor devices continued to make progress consistent with Moore's Law in the new millennium, despite the dire warnings of the late 1990s. As the industry entered the sub-100-nm regime, however, simple Dennard scaling failed to provide the same benefits enjoyed since the 1960s as devices became limited by materials challenges.

The industry responded by introducing a set of materials innovations that together came to be known as "equivalent scaling." The term "equivalent scaling" hints at the purpose of these materials innovations: to deliver performance benefits equivalent to those previously gained through Dennard Scaling. "Equivalent scaling" improves the variables from the transistor current-voltage relationship that did not scale (e.g. electron mobility, μ) or addresses quantum mechanical effects not accounted for in the basic current-voltage model (e.g. tunneling through the gate oxide) in order to continue scaling feature size. These innovations included the introduction of new materials (e.g. hi-K insulators, metal gates), strain engineering of the source and drain, and new device geometries (FinFET) coupled with continued reductions in transistor geometry. As engineers continued to tweak CMOS through material, process, and device

[39] These limits mimicked those reviewed in Keyes (1975).

improvements, the limits previously predicted by researchers were circumvented. Thus, the limits were to "CMOS as we know it" not CMOS per se.[40]

Despite the success of these materials innovations, CMOS would not remain indefinitely extendable. By the mid-2000s, the industry faced departures from its historical trajectory. Perhaps the most readily apparent of these departures was the abrupt plateau in microprocessor frequency improvements and the industry's switch to multi-core designs.[41] The 2003 ITRS forecast a nearly 18x increase in clock-frequency from 2003 (2.98 GhZ) to 2018 (53.2 GhZ). Instead, circuit-designers found themselves power-constrained and unable to match the historical rate of improvement in microprocessor frequencies. Needing to produce the next generation chip, firms introduced multi-core microprocessors for desktop computing. Gaining performance advantages with these multi-core microprocessors requires parallelization. However, only some tasks can be parallelized and doing so is expensive. Few programmers possess the skills required to execute effective parallelization (Thompson 2012). In summary, although equivalent scaling allowed semiconductor firms to improve transistors, net improvements in system performance failed to match their historical trends.

Achieving continued progress in semiconductors through equivalent scaling also grew increasingly expensive: as the technological challenges grew, semiconductor R&D expenditures throughout the 2000s as a percentage of sales were significantly higher than in decades past. With the onset of equivalent scaling the number of input knobs available to device engineers and circuit designers vastly increased, as did the research challenges. New challenges with materials

[40] Some researchers (aka Hoenesien and Mead, 1971) were explicit in laying out how their predictions were to "CMOS as we know it" but much of the popular discussion regarding the "end of Moore's Law" has been less nuanced.

[41] In mid-2004, Intel cancelled announced upgrades to its existing single-core desktop microprocessor line and began moving its product offerings toward multi-core designs ("Intel cancels Tejas" 2004)

integration and lithography limitations were accompanied by design constraints. These design constraints contributed to rapid increases in the cost to design and manufacture chips while also muting the performance benefits of scaling[42], worsening the economics of producing chips at the frontier of semiconductor process technology. Today, the ability to reduce cost is once again limited by lithographic technology: 193-nm lithographic techniques are still in-use at the 14/16nm production node through a combination of immersion lithography and new multi-patterning techniques. Firms are currently projecting continuing to use of 193-nm lithography in the upcoming 10 nm (late 2017) and possibly even 7nm (around 2020) product nodes. Extending existing lithography to those nodes, while technically feasible, may inhibit the industry's ability to achieve cost improvements as it complicates manufacturing and design flows. Beginning in 2013, firms at industry conferences began presenting data showing an increase in the cost-per-transistor at the newest process nodes (sub 28-nm).[43]

As a consequence of the compounding problems in device manufacturing and chip design, the cost of building new, leading edge manufacturing facilities has grown markedly – so much so that the industry has coined these rising costs "Moore's Second Law." Beginning in the early 2000s the worsening economics grew untenable for many vertically integrated firms. According to VLSI Research, the number of firms with integrated fabs at the process frontier began to dwindle, dropping from 29 in 2001 to 8 by 2015.. The industry's equipment suppliers underwent similar consolidation. Moreover, by the late 2000s, venture capital funding of new semiconductor firms also began to dry up. From 2007 through 2011 only 36 firms received initial

[42] For a brief discussion of how these challenges interact see Kahng, 2014, "Lithography-Induced Limits to Scaling of Design Quality."

[43] Others have noted that the right metric is "cost-per-function" which does show continual decreases but the increase in cost-per-transistor is a break in trend which lasted decades. Intel also claims its cost per transistor continues to decrease but that may be a function of its unique manufacturing processes.

venture capital funding, fewer than the number of firms (44) funded in 2003 alone (Lawler 2012). While this time period overlaps with the Great Recession and its aftermath, the rate of global startup formation in the industry appears to have began to decline as early as 2005 according to data compiled by Fontana & Malerba (2010)

5.2. The End of Moore's Law?

The industry's increasing difficulty in sustaining its historical trajectory has been further compounded by changes in its end markets. From the late 1970s through the early 2000s, the industry was driven by the personal computing market. However, by the mid-2000s, mobile phones - in particular smart phones - were the industry's fastest growing end-market. The standardized architecture, rapid growth, and short replacement cycles for mobile phones still offers incentives to invest in new fabs at the process frontier, primarily for the production of memory and logic chips. However, unlike in the personal computing market, smart phone manufacturers have captured the bulk of the profits in the smart phone market. Chips for smart phones, although developed with leading-edge process technologies similar to server and personal computer counterparts, command much lower prices. This maturation of traditional end-markets and growth opportunities in new sectors with lower margins further spurred M&A deals that consolidated the industry into a few leading players. [44]

The worsening economics of leading-edge semiconductor production, industry consolidation, and shifts in the industry's end-markets have also begun to fracture the industry's collaborative research ecosystem. During much of the 1980s and 1990s, the manufacturing ecosystem consisted of an array of leading-edge manufacturers serviced by an even larger group of equipment suppliers. In this context, the collaborative research ecosystem served as a

[44] https://www.pwc.com/us/en/technology/publications/assets/semiconductor-industry-device-deal-trends.pdf

coordinating mechanism for the vertically disintegrated ecosystem, transparently relaying technology needs throughout the entire industry. Firms producing chips in commodity markets (e.g. memory or CPUS) pushed investment in new process technologies and manufacturing equipment. Their investments benefitted the entire industry as the new process technologies and equipment quickly became available to fabless firms through foundries. This created a pool of research with large communal value, which industry, government, and academia collaborated together on through organizations such as Japan's VLSI, the SRC, SEMATECH and ITRS.[45] Throughout this period, firms benefited from large levels of R&D spillover given this common technology platform. For example, IBM's decades-long investment in basic research for the implementation of copper interconnects spilled over to the rest of the industry rapidly (Lim 2009). Within a few short years, other firms, many of whom had no history of basic research in copper interconnects, had introduced the technology into their production processes. According to Lim (2009), these knowledge spillovers were accelerated by the industry's shared knowledge base and collaborative research institutions facilitating technology transfer between firms.

However, industry consolidation has greatly reduced the number of firms interested in funding the type of pre-competitive research traditionally carried out by these organizations. As a result, the industry's long-standing collaborative organizations have seen reduced industry membership: Full membership in the SRC has been declining since the early 1990s and partial membership declining since the mid-1990s (see Fig 8 below). While manufacturing consolidation has eroded horizontal collaboration in pre-competitive research, consolidation amongst the industry's equipment suppliers has also affected the research organizations

[45] See Kapoor & McGrath (2014) for a discussion of how the mix of institutions presenting process technology research shifted as the technology matured.

coordinating vertical collaboration. In the last two years, SEMATECH has ceased to exist as a stand-alone organization, and its signature initiative, the Global 450 Consortium, stalled altogether after Intel and Samsung pulled out of the consortium. Similarly, while the early 2000s had several EUV consortia, in 2012 Intel, TSMC, and Samsung each took equity ownership in ASML, the leading EUV producer, to concentrate their R&D efforts (Intel Corporation 2012). Although the industry continues to rely on external R&D organizations, it has increasingly shifted to a customer-client model exemplified by Belgium's IMEC and so-called "customization" programs at SRC. These programs allow member firms to identify specific projects of interest at the R&D organizations and to earmark their funds for that work. In effect, a portion of the collaborative R&D organization has become a private contract organization. While firms still garner benefits of external research, this structure reduces the public-goods aspect of the research. The 2013 edition of the ITRS was the last edition to be supported by SEMATECH and the SRC, and the 2015 ITRS edition, released in July 2016, was the last to be sponsored by the SIA.

This combination of numerous challenges - worsening economics of production, maturing end-markets, industry consolidation, and fracturing of the industry's pre-competitive collaborative research structure comes as the industry faces its greatest technological uncertainty in decades. In the late 1990s and continuing through the 2000s, industry researchers began mapping out divergent trajectories for the CMOS platform. In the 1994 NTRS, these paths were labeled "evolutionary" and "revolutionary" changes to CMOS. By 1999, however, the industry had begun to more specifically bifurcate the trajectories. The SRC began two small research programs in 1999 to fund research to both "reach the limits of CMOS technology" but also "provide the basis for inventing new technologies that will eventually supplant CMOS" (Cavin

III et al. 2000). By the late-2000s, as CMOS had begun to evolve with equivalent scaling technologies, the ITRS adopted a three trajectory typology. The ITRS coined these three trajectories "More Moore" (i.e. continuing the historical trajectory of performance improvements with continued CMOS evolution), "More-than-Moore" (heterogeneous functionality integration into the CMOS platform), and "Beyond CMOS" (everything from a new computing element to entirely new computing architectures that are at least initially CMOS compatible).

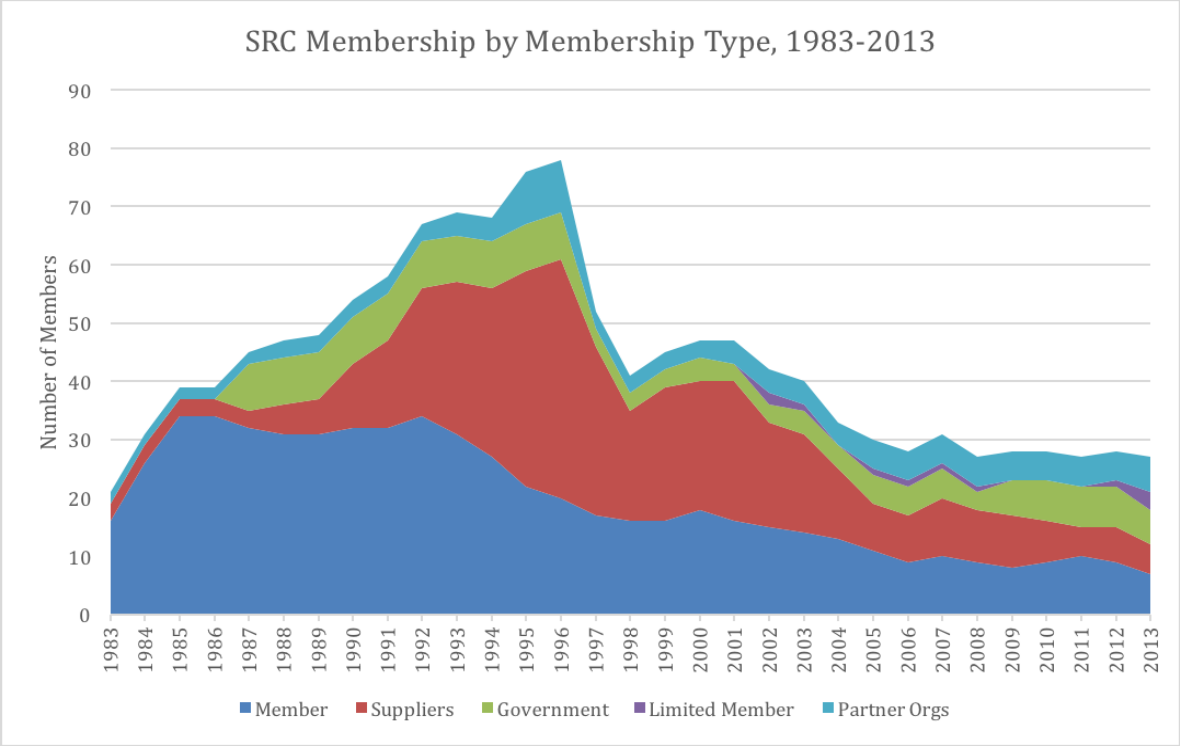

Figure 8 – SRC membership over time. Data from SRC records.

Under the ITRS' current schema, CMOS remains central to computing technology for the foreseeable future. Unlike the industry's single trajectory of the past few decades, the industry envisions several concurrent areas of advancing technology and an eventual transition to an entirely new technology, "Beyond CMOS." Advances in More Moore, which are pursued by only a few firms that compete in markets for commodity products (e.g. DRAM, Flash, CPU), continue to form the core of semiconductor technology to be supplemented by More-than-Moore

and “Beyond-CMOS” technologies. “More-than-Moore” technologies are customized to specific applications, and may have a longer timeline than More Moore, but lack the common R&D platform that drove Moore’s Law. Meanwhile, “Beyond-CMOS” technologies are “bottom-up”, requiring fundamental advances in basic science and initially foreseen to be CMOS compatible with niche applications. In the long-term, a heretofore unknown “Beyond CMOS” technology (or combination of technologies) is envisioned as offering continued extendibility of computing advances. However, as already noted, that requires the commercialization of new science and is further complicated by the interdependencies of integrating a radically new computing element within the existing information technology hierarchy.

More Moore is, true to its name, a continuation of the industry's historical trajectory of improvement in power (energy per switching event), performance (operating frequency), area (density) and cost for transistors. According to the ITRS, the primary technical drivers for continued advancements through 2020 will be similar to those of the past decade: implementation of new device geometries (e.g., gate-all-around transistors and nanowires), integration of new materials for the transistor channel and interconnects, and a possible switch to tunnel FETs beginning sometime after 2020 (ITRS 2016). While these projected evolutions of CMOS through 2020 are further refinements of the equivalent scaling techniques in use since the late 1990s, the scope of technical challenges facing integrated circuits goes beyond those addressable by device-level innovation. For example, transistor leakage current and interconnect delay across a chip have continued to worsen, and commercially available transistors are close to the physical limits for subthreshold slope.[46] Leakage power has grown to become a substantial

[46] Intel claims its 14nm process is around 65 mV/decade, close to the 60 mV/decade fundamental limit (Natarajan et al. 2014).

portion of total power consumption, and power consumption has limited the benefits of further scaling (Horowitz 2014). As a result, computing is power-constrained with consequences for system and architecture design (Huang et al. 2011), and even programming. Furthermore, as noted above, the increasing complexity of designing leading-edge chips and delays in integrating next-generation EUV lithography have contributed to a slowing of improvements in cost-per-transistor at the newest nodes. For these reasons, many industry experts have argued that continued transistor scaling delivers only muted performance and density improvements relative to the historical trend. Matching these increasing costs to the reduced revenues of these microprocessors as they become low-value commodities in mobile markets, suggests this trajectory is likely to have limited life.

The second path forward has been dubbed "More-Than-Moore" and is application driven. European researchers first put forth a definition for More-than-Moore in the late 1990s, "functional diversification of semiconductor-based devices." This trajectory was first included in the 2011 version of the ITRS. Our discussion of this trajectory includes both the ITRS definition but also a new trend of firms exploring an array of alternative methods to improve performance for their specific applications as the relative improvement in power, performance, cost and area from More Moore techniques has reduced. As a consequence, in recent years foundries have noted an increase in the lifespan of "legacy" or older process nodes as firms attempt to wring out increased performance from older process technology.[47] Additionally, end-users have begun to look beyond standardized products for solutions, instead employing more specialized chips tailored to specific applications. For the purposes of this paper, we group these chips tailored to

[47] Data compiled from TSMC's earning reports shows long-lived technologies such as the 0.15 and 0.18 micron processes introduced around the turn of the century still account for 10% of sales in 3Q 2016 (Culpan 2016).

specific applications into the category of "More than Moore." Examples include the use of FPGA's to power a data center by search companies (e.g. Microsoft Bing) in lieu of CPUs, the increasing use of GPUs for machine learning tasks, and Google's announcement of an ASIC developed in-house for its deep-learning activities. To lead these developments firms like Apple (2008), Amazon (2015), and Google have all brought chip design in-house. These trends re-shape the industry's traditional market structure and supplier-customer relationships. As large end-users vertically integrate chip-design capabilities, they begin to compete with their current suppliers. Further, the benefits of R&D in these specialized applications are limited to firms with those applications, in contrast to the industry-wide benefits of advances in the underlying transistor technology. Finally, while these approaches have introduced real benefits for firms, their extendibility may still rely on advances in underlying transistor technology that require the sort of basic research conducted almost exclusively by the semiconductor industry's largest manufacturers.

Under the ITRS' definition for More-than-Moore, a new area of research is the integration of entirely new functionalities into a system including novel sensors, radio-frequency (RF) or micro-electromechanical (MEMS) devices. Integrating these capabilities requires the development of new manufacturing processes, design methodologies, and possibly business models. These technologies differ from previous advances in semiconductor and computing because they are tied more directly to applications with specialized needs. As a result, they require collaboration across industries (e.g. medicine and semiconductor manufacturing) but, like the firm-driven application-focused research, also lack the shared R&D platform which helped drive progress in microelectronics. Unlike the industry's historical roadmaps which projected

out existing trends and attempted to identify roadblocks, the technical needs of these applications are shaped by a litany of non-technical factors.

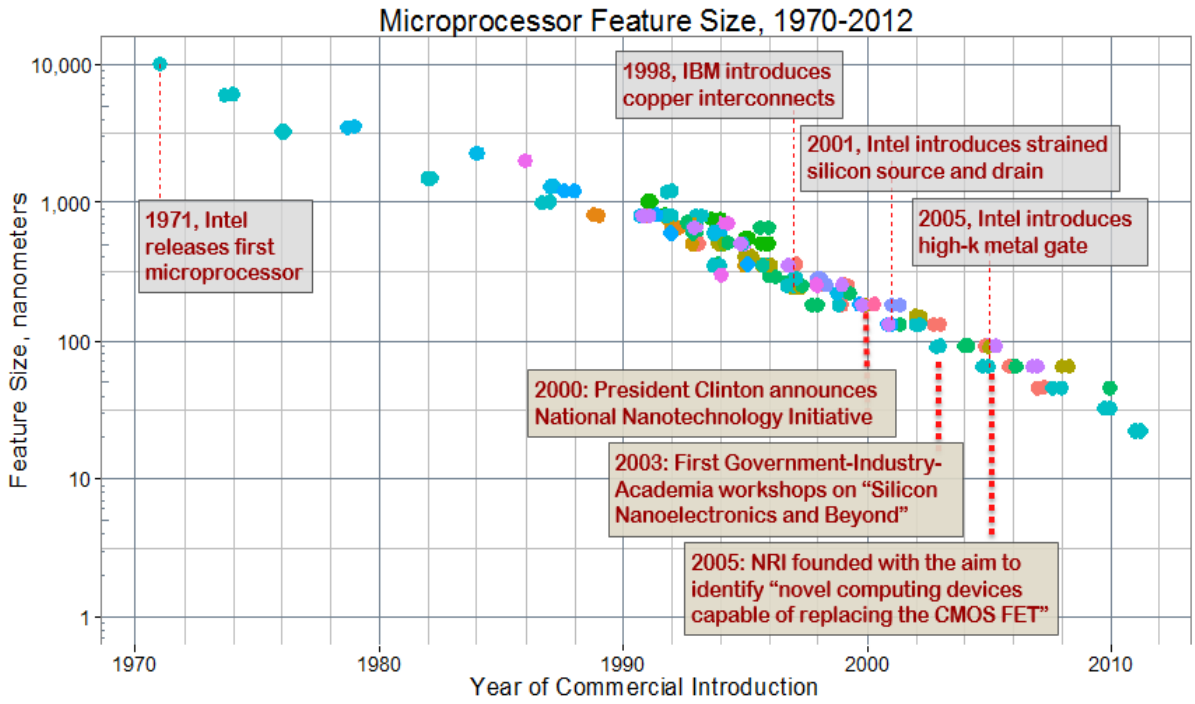

Figure 9 - Materials innovations keep scaling alive through the 2000s; new institutions to address long-term technology challenges. Adapted from Danowitz et al., 2012.

While firms are already competing in both the More Moore and More-than-Moore trajectories, the most uncertainty and perhaps the greatest potential is in the "Beyond CMOS" space. Despite the industry's nearly 50-year history of specious Chicken Littles, today's fundamental limits to the improvement of transistors are being taken seriously by a small but important set of industry and academic researchers. Beginning with the 2001 roadmap, the ITRS included a new chapter, "Emerging Research Devices," which speculated on the evolutions (i.e. "non-classical CMOS") needed for CMOS to reach the "end of the Roadmap timeframe" and possible technological revolutions that would extend microelectronics "beyond the end of the Roadmap", i.e. beyond the roadmap's 15 year timeframe (ITRS 2001). The chapter's final two sections on Emerging Logic and Architecture were motivated by the prediction that "the scaling

of CMOS device and process technology, as it is known today, will end by the 22 nm node (9 nm physical channel length) by 2016" (ITRS 2001). The chapter's Emerging Logic section included a bevy of potential successors to the silicon CMOS transistor: single electron transistors, resonant tunneling diodes, and molecular devices, amongst others. In the Emerging Architecture section, the authors noted that the new device structures highlighted in the Emerging Logic section "will require new architectures to achieve useful functionality." In other words, extending computing beyond the capabilities of CMOS required the introduction of new computing elements and computing architectures. The chapter concluded with an early quantitative estimate of the capabilities of emerging technologies relative to ultimately scaled CMOS, with the authors concluding "few of the new technologies are directly competitive with scaled CMOS and most are highly complimentary [sic]" (ITRS 2001). Many of the concepts surveyed in the initial chapter had research lineages that dated back to the industry's then-defunct basic research laboratories and kept alive by military agency funding through the 1990s and early 2000s.[48] While the 2001 chapter noted that many of the devices were "speculative" and not directly competitive with CMOS, by 2003 the ITRS placed these devices as central to the future of the industry. The report's executive summary highlighted the ERD chapter's discussions of post-CMOS devices as "pav[ing] the way to a complete technological revolution looming ahead towards the end of the next decade" (ITRS 2003). Moreover, the 2003 ERD chapter was built around yet another limit, an analysis by SRC researchers which concluded that, "even if entirely different electron transport devices are invented for digital logic, their scaling

[48] Concepts such as RSFQ could in part be traced back to IBM's Josephson Junction program of the 1970s. TI's Central Research Laboratory conducted early work on "tunnelling devices" that was extended with DARPA's ULTRA (1990-1998) program. DARPA programs throughout this period laid foundations for emerging devices: Spintronics (1994-2000), QuIST, SPINS, and Moletronics programs. The SRC also began a modest program of funding "novel devices" in 1999.

for density and performance may not go much beyond the ultimate limits obtainable with CMOS technology, due primarily to limits on heat removal capacity" (Zhirnov et al. 2003).[49] Building on the conclusion of the above analysis, the ERD report concluded that none of the existing research devices were "viable emerging logic technologies for integration" (ITRS 2003).

Table 1 - Summary of technology inclusion in ITRS Emerging Research Devices Chapters, 2001-2013

TECHNOLOGY/ YEAR	2001	2003	2005	2007	2009	2011/2013
RSFQ	X	X				
1D STRUCTURES		X	X			
RESONANT TUNNEL	X	X	X			
SINGLE ELECTRON	X	X	X	X	X	
QCA	X	X				
SPIN FET		X	X	X	X	X
SPIN VALVE		X	X			
FERROMAGNETIC LOGIC			X	X		
TUNNEL FET					X	X
IMPACT IONIZATION					X	X
NEMS SWITCH					X	X
NEGATIVE GATE CAPACITANCE FET					X	X
MOVING DOMAIN WALL					X	
NANOMAGNETIC LOGIC					X	X
MOTT TRANSISTOR						X
SPIN WAVE					X	X
EXCITONIC FIELD EFFECT						X
SPIN TORQUE MAJORITY LOGIC GATE						X
ALL SPIN LOGIC						X
MOLECULAR ELECTRONICS	X	X		X	X	
ATOMIC SWITCH					X	X
BISFET					X	X

[49] A few years later researchers from IBM concluded that a different mechanism, source-to-drain tunneling (Haensch, E.J. Nowak, et al. 2006) would serve as the ultimate limit to device scaling at 7nm rather than Zhirnov et al's 3nm.

This Chicken Little elicited action among industry players in the barnyard, although the action was paltry at best. Over the course of the next two years, industry leaders, acting under the aegis of the SIA, worked to establish a "nanotechnology strategy" to address the industry's long-term research challenges and leverage vast new public funds for nanotechnology research allocated by the 2001 National Nanotechnology Initiative. The culmination of this work was the establishment of a new consortium, the Nanoelectronics Research Initiative, co-funded by 6 US semiconductor manufacturers[50] with additional public funding from both federal and state sources. Total combined funding through both public and private sources has averaged approximately $20 million per year since the NRI's founding in 2005. In creating the NRI, the industry aimed to explicitly address the lack of viable alternatives highlighted by the ITRS. The NRI's stated goal at founding was "demonstrat[ing] novel computing devices capable of replacing the CMOS FET as a logic switch in the 2020 timeframe" (SRC 2006a). In 2013, the program was re-competed and the NRI refined its goal to "[d]emonstrate non-convential, low-energy technologies which can outperform CMOS on critical applications in ten years and beyond" (SRC 2013). In 2013, the SRC re-competed its FCRP program first established in 1997. The new program, renamed STARnet included funding for three centers investigating "Beyond CMOS" devices and materials. Total funding for STARnet is approximately $50 million annually for a total of six centers. In the US, the SRC programs are the leading organized efforts investigating beyond CMOS technologies, with additional support for basic science coming from federal research agencies. While a survey of nanoelectronics funding in the US, Europe, and Japan found that funding totals for critical research areas in nanoelectronics had improved since 2009 (Galatsis et al. 2015), the bigger picture is that comparatively fewer resources are being

[50] AMD, Freescale, IBM, Intel, Micron, and Texas Instruments. Only Intel, IBM, Texas Instruments, and Micron remain members today.

devoted to invent, develop and commercialize beyond CMOS technologies than were devoted to previous industry-government partnerships such as the $200 million spent annually on SEMATECH from 1987 through 1997.

Despite the industry's relatively modest funding of NRI and related efforts, the Beyond CMOS trajectory presents the greatest scientific and technological uncertainty. This uncertainty is evident in the ITRS' Emerging Research Device (ERD) chapter. Since its inception in 2001, the ERD chapter has included over 20 different proposed logic technologies with entries removed or added based on the progress made in the intervals between publication. The inclusion of technologies in the ERD from 2001 through 2013 is summarized in the table above. The current ERD typology for logic devices is reproduced below with "Beyond CMOS" devices confined to the two occupied boxes on the right. According to the ITRS, a CMOS successor "is thought potentially to be viable" if it offers advantages in one or more of the following: device density and overall cost, switching speed, switching energy and overall energy, or "enabling of novel information processing functions" (ITRS 2016). It is only in recent years that researchers, first under the aegis of the NRI and now as a joint program between the NRI and STARnet, have begun to develop standardized benchmarking programs for the comparison of proposed devices to CMOS. These early benchmarks have attempted to compare the operation of the proposed technologies to CMOS in standard circuit configurations.[51] However, even here researchers are hamstrung by the fact that many of the technologies lack working prototypes. Furthermore, these benchmarks attempt to compare devices in their performance at tasks for which CMOS is well-suited, potentially underplaying the value of new devices that may lack in traditional metrics

[51] Bernstein et al (2010) and an updated set of benchmarks were released in late 2015 led by Intel researchers in coordination with the NRI (Nikonov, 2015)

(e.g. switching speed) but offer entirely new functionality (e.g. non-volatile operation). At the same time, the inter-dependencies between physics, material, device, and architectural advances make such decisions difficult to evaluate and benchmark against, since the possibilities in many of the other interdependent components remain unknown. Developing the methodology to properly compare and evaluate CMOS-based systems against systems designed from the ground-up around new device technologies remains out of reach for researchers.

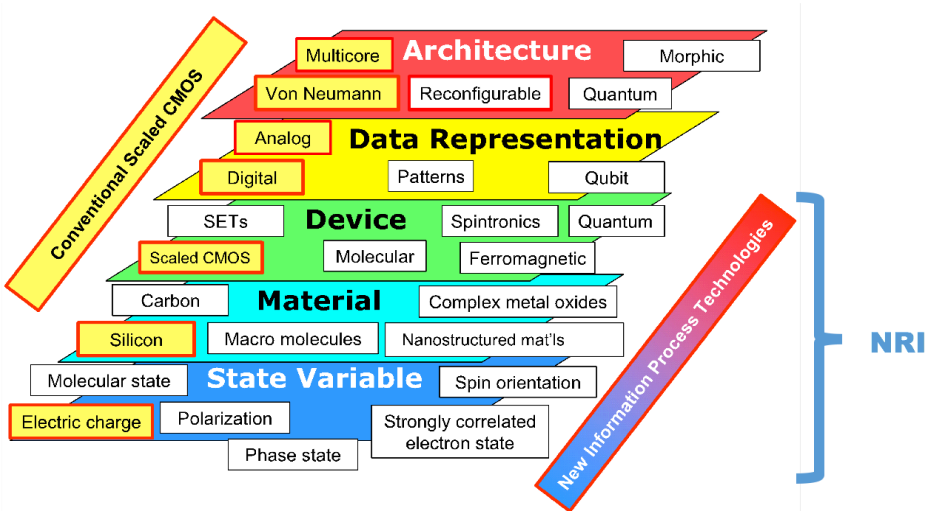

Figure 10 - Nanoinformation Processing Taxonomy Source: Emerging Research Devices (ITRS 2007).

As documented above, the industry's past is replete with the failed development of ambitious alternatives to the silicon integrated-circuit. A commercially successful alternative will require reducing to practice an entirely new computing element based on a new materials system that likely operates using a different set of physical phenomena than the industry has traditionally relied on. The interdependency of these challenges has been recognized by the ITRS and is depicted below. The industry's insistence on any new technology's compatibility with CMOS serves to reduce complexity but also adds constraints that may hamper the development of any promising new technology. Beyond the challenges of scientific discovery and invention of new

device concepts, the commercial success of such a technology will likely require the development of new manufacturing techniques and design tools to bring these devices to market. The concomitant investments in equipment and education represent looming financial challenges to an industry grappling with the worsening economics of its current product offerings.

Thus, unlike any time previous in the semiconductor industry, firms face a bewildering number of potential options going forward with incredible levels of uncertainty over which technology to back. The industry's current predicament echoes the period of the late 1950s, prior to the invention of the integrated circuit when researchers pursued a bevy of solutions to the "tyranny of numbers." Today, however, the industry finds itself with several distinct technology challenges. Maintaining the historical gains in the silicon CMOS paradigm, extending the capabilities of CMOS with new device functions, and inventing and commercializing a successor technology that once again offers a platform for advancement in the decades to come.

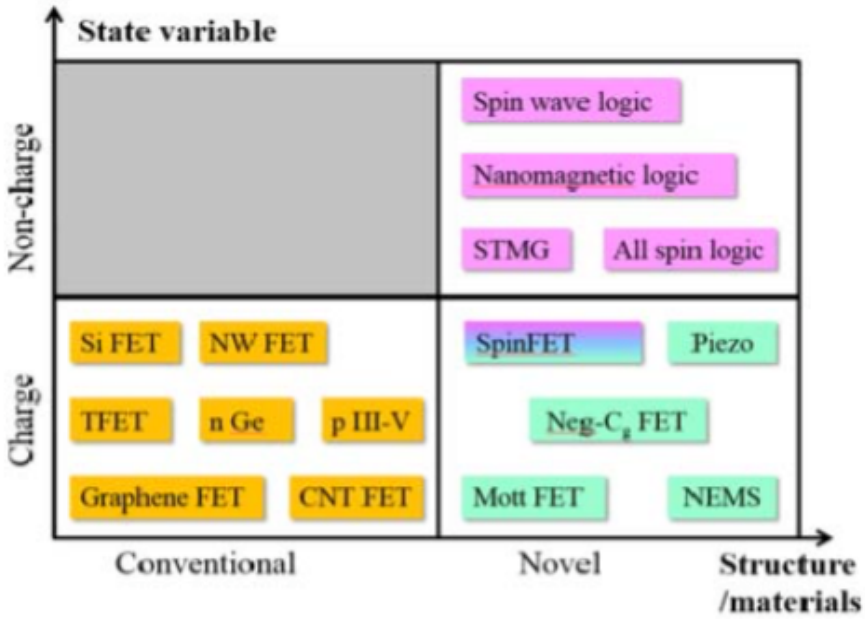

Figure 11 - ITRS 2.0 Emerging Research Device taxonomy. Source: ITRS, 2016

6. Conclusions

For over 50 years, the semiconductor industry has successfully brushed off the warnings of various Chicken Littles. The extension of the silicon integrated circuit paradigm from the 1960s through present day has been made possible by a combination of technological breakthroughs to overcome proposed limits and organizational responses to a changing competitive and industrial landscape. Rising R&D and manufacturing costs were offset by the growth in demand made possible by exponential improvements in the cost, speed, and energy efficiency of transistors. Decades of cumulative improvements democratized computing, moving it from a handful of room-sized behemoths to billions of pockets. Moore's Law became a self-fulfilling prophecy, industry benchmark, and key driver of global productivity growth.

Despite this, for many of the barnyard animals, the semiconductor industry's travails today do not seem worth raising alarm. Many have perhaps been lulled into a sense of complacency by the litany of "failed" Chicken Littles throughout the industry's history. Some observers are optimists and point to the numerous scientific advances in materials and devices reported by researchers in the last few years and their concomitant faith in engineers' ingenuity to invent around limits. Past performance is an indicator of future success. Other members of the barnyard are not optimists but simply disinterested. To them the semiconductor industry today represents a maturing industry, whose founders were the trail-blazing capitalists of yester-year. From this vantage point, the slowing market growth and increasingly difficult technology advances are par for the course; silicon integrated circuit technology is simply approaching the end of its S-curve and other technologies are poised to assume the role semiconductors played in the latter half of the 20th century onward.

Absent a radical change, the future of computing is increasingly fragmented, with potentially large economic and social consequences. As the economics of the semiconductor industry's historical trajectory have worsened, the industry's participants – both traditional manufacturers and the customers who have newly vertically integrated – are pursuing divergent technological paths. This dynamic threatens to undermine the contributions from semiconductors as a general purpose technology. Firms attempting to leverage advances in a general purpose technology need to know what it will be capable of in the future (Bresnahan and Trajtenberg, 1995). "Moore's Law" embodied the relatively stable trajectory in integrated-circuit technology over the last 50 years. However, as advances in semiconductors slow and downstream firms increasingly pursue application or domain specific innovations, technological progress will be increasingly unevenly distributed.

While it is undoubtedly true that there has been considerable progress in related fields – e.g. computer vision and artificial intelligence – these advances do not diminish the need for faster, cheaper, and more energy-efficient computation. Consider the most bandied about visions for future technological capabilities: traffic-free streets navigated by driverless cars, error-free medical diagnoses by advanced artificial intelligence systems, widely-dispersed intelligent sensor networks that reshape our relationship with the world around us, and scientific breakthroughs enabled by advanced models of complex systems. Each of these endeavors will require domain-specific advances but central to each is the need for faster, more energy-efficient, and cheaper computation. Finally, while these divergent paths may extend advances in computing performance in niche applications for another decade or three, the underlying problem of how to advance computing when these temporary fixes run out will remain. While industry interests in extending the existing S-curve through niche application fixes may be

socially beneficial and profitable, they will draw attention and funds away from the core problem: a solution to advancing computing beyond CMOS. This network externality – both the social and economic benefits of continued advance in computation being underserved by private incentives – is exactly the type in which government funding is essential. Although continued evolution of CMOS technology is proving to be expensive and un-economic, the extendibility of a possible successor technology may offer the best chance at relatively widespread social and economic gains.

As it stands today, however, the response from industry, academia, and government has been lacking. In terms of public funding, aside from the relatively small amounts of funding devoted to basic research, mission oriented agencies are driving the evolution of emerging computing paradigms. The Departments of Energy and Defense are working closely with researchers on developing computing technologies for the needs of their respective communities. Yet, it is unlikely that the solutions devised for these niche applications, at least currently, will broadly spillover to general computing technology in the near term. For example, the early benefits of quantum computing are limited to specific classes of algorithms and some of the technologies being considered for high-performance scientific computing require operation at supercool temperatures, a technology unlikely to make its way into the homes and pockets of the average consumer.

Unfortunately, while the industrial, economic, and social consequences are tremendous, the combination of technological, economic, and organizational challenges facing the industry today are unprecedented, and unlike those faced at any point in the industry's past. Today, the industry's institutions that historically shaped the evolution of basic science into technology are weakened. Central research labs and military demand, both of which played instrumental roles in

the industry's earliest periods, have either dried up, in the case of the former, or no longer are the primary source of demand, in the latter. Similarly, the collaborative institutions that coordinated the industry's vertically disintegrated research ecosystem beginning in the early 1990s have been fractured. Lack of outside investment and maturing product markets have driven firms to consolidate, reducing the players and funds for collaborative institutions. Increasing focus on application-specific advancements has further reduced the extent to which technological advancements benefit all and thus further reduced the incentives for collaborative R&D. Instead, the industry's customers are vertically integrating to incorporate firm-specific semiconductor capabilities. The industry's collaborative research enterprise is thus facing a secular trend toward decreasing pre-competitive funding as the remaining traditional firms begin to earmark their external research funding toward specific projects and technologies. The slowing pace of advancement along the industry's historical trajectory also raises the specter of commodification. China in particular has been investing heavily to catch-up to the technological frontier, which may be easier if the pace of advancement slows.

Re-establishing the extendibility of semiconductors will likely require the commercialization of technologies based on heretofore undiscovered scientific breakthroughs. As highlighted in the move from single to multi-core processors, downstream technology shifts require coordination up and down the computing technology stack to be fully utilized. However, the current environment lacks the sort of coordinating institutions that guided previous technology shifts (e.g. large corporate research laboratories in the case of vacuum tubes to transistors) – and, the collaborative R&D institutions that remain, are being broken apart. With limited investment in basic research and firms pursuing divergent technology tracks, the sky truly may be falling, with citizens world-wide to pay the economic and social consequences.

Scaling Moore's Wall: Institional Responses at the End of a Technology Paradgm

1. Introduction

Over the last several decades, the structure of industrial R&D has undergone considerable change with a shift away from centralized R&D operations, an increasing reliance on external sources of innovation, and reduced investment in basic science (H. W. Chesbrough 2003; Mowery 2009; Arora et al. 2015). One consequence of the changes to the nature of industrial R&D processes is that during long periods of incremental change firms are often unwilling or unable to maintain the "systems level knowledge" necessary to carry out—or even adapt to—radical architectural shifts in technology (Chesbrough and Kusunoki, 2001). Although there is considerable academic literature on the mechanisms by which firms can absorb and influence externally occurring innovations (Cohen & Levinthal 1990; Powell et al. 1996; Ancona & Bresman 2007; Rosenkopf & Almeida 2003; Mowery & Teece 1996), there is a relatively less work examining the ability of institutions to address technological discontinuities when the underlying scienceto navigate through that discontinuity is unknown.

Our study examines the emergence of the Nanoelectronics Research Initiative (NRI) as the technological community's response to a looming discontinuity: the end of Moore's Law. We show how a key group of industry and academic leaders created consensus on the need for research into new device technology, despite the long history of failed research into alternatives to the silicon integrated circuit. Advancements in semiconductors were the single largest contributor to total factor productivity growth during the 1990s (Jorgenson 2001); therefore, the

economic value of scaling Moore's Wall is tremendous and goes well beyond the pecuniary benefits of chip manufacturers. In order to continue apace with the historical cadence of Moore's Law the semiconductor industry must identify a new computing device, or switch, capable of replacing the complementary metal-oxide-semiconductor (CMOS) field-effect transistor (FET) that has prevailed as the dominant technology for more than 30 years. The discovery and development of a new switch on a new (non-CMOS) material platform requires investments in basic research, which today has largely moved out of the industry's corporate labs. A new switch in a new material system also threatens to de-value the core competencies developed by firms across the industry's supply chain – from chip designers and equipment producers to software programmers and semiconductor manufacturers themselves. As community consensus began to emerge around this challenge, leading U.S. semiconductor firms established the Nanoelectronics Research Initiative (NRI) in 2005 with the stated mission of "demonstrat[ing] novel computing devices that could enable the industry to extend Moore's Law improvements in electronics far beyond the limits of CMOS" (SRC 2005).. The NRI was the second public-private partnership formed within the Semiconductor Research Corporation (SRC) since its establishment by the U.S. Semiconductor Industry Association (SIA) in 1982. It represents a departure from both the SRC's traditional funding and research models. The NRI is the first SRC program funding research exclusively outside of the silicon CMOS paradigm and did so primarily through public funding.

Previous research examining the role of institutions in addressing technological uncertainty has not examined cases of paradigm change in open systems (Tushman & Rosenkopf 1992). Past literature has focused on public-private partnerships aimed at aiding national industries that had fallen behind international competition. Prominent examples from the

semiconductor industry, Japan's Very Large Scale Integrated Circuit (VLSI) program (1981-1984) and the United States' SEMATECH program (1987-1996), focused initially on supporting industry catch-up resulting especially in improvements in the capabilities of domestic suppliers (Sakakibara 1993; Spencer & Grindley 1993). Furthermore, while other research has argued that conferences (Garud & Rappa 1994; Garud 2008), or communities of experts (Haas 1989; Rosenkopf & Tushman 1998) can play important roles in technology direction-setting, these occurred within a technology paradigm. Our research focuses on a public-private partnership established with the goal of long-term scientific discovery and early technology development. We look specifically at an industry facing a technological discontinuity where the solution and supporting science are almost wholly unknown.

This chapter triangulates 50 oral histories and semi-structured interviews, 70 hours of participant observation, public news outlets, and extensive access to the internal archives of the SRC (Jick 1979). In the tradition of (Rosenberg 1976; Henderson 1995; Bates et al. 1998; Levi 2002; Ingram et al. 2012; Kahl et al. 2012), we develop an analytical narrative. In developing this narrative, we do not seek to build grounded theory (e.g. Glaser & Strauss 2012; Eisenhardt 1989), rather we believe that the emergence and existence of the NRI in response to the end of a technological paradigm is of itself interesting to existing theory. We seek to examine the public private partnership model presented by the NRI and to analyze the processes its members used to achieve a major scientific and technological breakthrough in a vertically-disintegrated industry. However, given that the NRI program was ongoing as of the terminal date of our study, we make no normative judgment about NRI's success or failure in meeting its objectives.

Our research suggests that a core group of industry and academic researchers were instrumental in creating the consensus on the need for a research program dedicated to finding an

alternative to the CMOS FET. Furthermore, we find that many of these same individuals were instrumental in coordinating the research directions of the wider scientific community, beyond the boundaries of the NRI. Specifically, we show how, building on a long history of managing industry-university research programs within the Semiconductor Research Corporation, NRI incorporated industry expertise in manufacturing and design to inform and shape academic research, all with the explicit objective of finding an equally fecund alternative switch to the CMOS FET. We conclude by questioning the extent to which the current effort is appropriately suited to the size and nature of the scientific, technical, economic, and social challenges posed by Moore's Wall. In particular, we note that overall public funding levels in the United States in the field has been largely stagnant since the founding of the NRI. We close by discussing why each institution – markets, government, and collaborative institutions – have failed to raise the seemingly requisite resources to address the challenge of Scaling Moore's Wall.

2. Methods and Data

This paper triangulates 50 oral histories and semi-structured interviews, 70 hours of participant observation, public news outlets, and extensive access to the internal archives of the SRC (Jick 1979). In the tradition of (Rosenberg 1976; Henderson 1995; Bates et al. 1998; Levi 2002; Ingram et al. 2012; Kahl et al. 2012), we develop an analytical narrative. In developing this narrative, we do not seek to build grounded theory (e.g. Glaser & Strauss 2012; Eisenhardt 1989) rather we believe that the emergence and existence of the NRI in response to the end of a technological paradigm is of itself interesting to existing theory. This analysis draws from three types of data: archival sources, semi-structured interviews, and participant observations (Jick, 1979). To date, we have conducted 50 interviews of member company researchers and executives, NRI researchers and center leaders, and SRC representatives to understand how those

in the NRI envisage and assess its operations (see table 1 below). We have also conducted oral histories of individuals who played prominent roles in the formation and evolution of the industry's key institutions including SRC, SEMATECH, and NTRS. Their first-hand perspective complement the archival records generated by these evolving institutions.

Table 2 - Archival Data Sources Used for Chapter 2.

Archival Data	Nanoelectronics Research Initiative (NRI)	Parallel Activities Outside NRI
Inputs	Firm financial data (Standard and Poor's Compustat), primary source planning documents from NRI organizers	Government funding data: NSF funding archives, DARPA public funding archives (NSF, 2013; RDDS, 2013)
Institutional	NRI annual review presentation archives (2006-2013), SRC annual reports (1982-2013)	Key industry and government conferences (IEDM, DAC), Industry publications (EETimes, ACM)
Individuals	NRI executive and technical board representation, NRI conference attendees	Program directors at federal research funding agencies (ONR, DARPA, NSF, NIST)
Outputs	NRI research publications and patent filings	Patents and publications in NRI topic areas (U.S.PTO, 2013)

We conducted some 70 hours of participant observations of SRC and NRI conferences and annual research reviews to observe firsthand the organization's deliberative processes. We attended the SRC's Techcon Conference in 2012 and the NRI's Annual Reviews in 2012, 2013 and 2014. We also observed NRI Technical Program Group conference calls in the summer of 2013, during which NRI industry representatives discussed NRI administrative issues. SRC's Techcon Conference provided an opportunity for all SRC-funded students to present their research to industry members and receive technical feedback. At the NRI annual review, academic researchers presented research updates to industry and government representatives. Attending these conferences allowed us to witness firsthand how these programs operate and form our own assessments.

SRC granted us considerable access to its archives, allowing us to analyze the historical development of SRC and providing historical context for the development and mission of NRI. Archival data include SRC newsletters and public announcements and access to SRC's personnel database and research reports. SRC also maintains a database of all research contracts and industry and academic personnel involved with SRC. These databases were used to identify key individuals in conjunction with "snowball" methods based on recommendations from other interviewees. Archival data utilized for this paper are listed in Table 2; a full listing of SRC-provided archival data appears in Appendix A.

Table 3 - Full list of interviews completed

	Name	**Organization**	**Role**
Semiconductor Firms	Tom Theis	IBM/NRI	Director of NRI
	Jeff Welser	IBM/NRI	Ex-Director of NRI
	Paolo Gargini	Intel	Ex-Senior Fellow (Retired)
	George Scalise	SIA	Ex-President (Retired)
	Bob Doering	Texas Instruments	Senior Fellow
	George Bourianoff	Intel	NRI TPG
	Dimitri Nikonov	Intel	NRI TPG
	Steve Kramer	Micron	NRI TPG
	An Chen	GlobalFoundries	NRI TPG
	Luigi Colombo	TI	NRI TPG
	Wilfried Haensch	IBM	NRI TPG
	Ajay Jacob	GlobalFoundries	NRI TPG
	Zoran Krivokapic	GlobalFoundries	NRI Governing Council member
	Tony Low	IBM/Purdue	NRI Student/NRI Liaison
	Tak Ning	IBM	NRI Liaison
SRC	Larry Sumney	SRC	CEO of SRC
	Ralph Cavin	SRC	Chief Scientist (retired)
	Victor Zhirnov	SRC/ITRS	Director, Cross-Discplinary Research
	Steve Hillenius	SRC/GRC	Director, GRC
	Gil Vandentop	SRC/STARNet	Director, STARNET
	Celia Merzbacher	SRC	VP of Innovative Partnerships

	Mihail Roco	NSF/NNI	Founder of NNI
Government	Lawrence Goldberg	NSF	Senior Engineering Advisor
	Clifford Lau	DoD/ONR	DoD NNI Representative
	Kerry Bernstein	DARPA	ex-NRI Liaison, Program Manager MTO
	Larry Cooper	ONR	Retired Head of Nanoelectronics
	Chagaan Bataar	ONR	Program Officer Nanoscale Electronics
	David Seiler	NIST	Chief of Semiconductor Division
	Curt Richter	NIST	Head of Nanoelectronics Group
	Yaw Obeng	NIST	Senior Scientist
NRI Center Directors	Kos Galastis	UCLA	Co-Director, WIN
	Alan Seabaugh	Notre Dame	Director, MIND
	Robert Dunn	Notre Dame	Co-Director, MIND
	Alain Diebold	SUNY Albany	COO INDEX
	Evgeny Tsymbal	Nebraska-Lincoln	Director, CNFD
Other University Researchers	Sayeeh Salahuddin	Berkeley	WIN Researcher, Former NRI Student
	Jeremy Levy	Pittsburgh	NRI-NSF Funded Researcher
	Randall Feenstra	CMU	STARnet (LEAST) Funded Researcher
	Patrick Fay	Notre Dame	NRI and STARNET Funded Researcher
	Susan Fullerton	Notre Dame	NRI and STARNET Funded Researcher
	Grace Xing	Notre Dame	NRI and STARNET Funded Researcher
	Kirill Belashchenko	Nebraska-Lincoln	NRI Funded Researcher
	Alexei Gruverman	Nebraska-Lincoln	NRI Funded Researcher
	Xia Hong	Nebraska-Lincoln	NRI Funded Researcher
	Christian Binek	Nebraska-Lincoln	NRI Funded Researcher
	Peter Dowben	Nebraska-Lincoln	NRI Funded Researcher
	Uttam Singisetti	Univ. of Buffalo	NRI Funded Researcher

3. Building on Existing Institutions: Emergence of a Public Private Partnership in Response to a Presumptive Anomaly

3.1. Anticipating Moore's Wall

As early as 1958 engineers at Westinghouse proposed a radical vision, "molecular electronics," to officials in the Air Force as a way to leapfrog the reliability limitations of interconnecting discrete silicon transistors (Holbrook 1995; Choi & Mody 2009). Although by early to mid-1968 the program, beset by technical complications and failed deliverables due to the lack of atomic-level control of materials necessary to succeed, migrated toward a more conventional integrated circuit approach, it represented the first in a series of technologies

proposed by the semiconductor industry, which were inspired by an understanding of future conditions under which the existing silicon system would fail, but that failed to supplant silicon.

The rapid, unrelenting extendiblity of silicon integrated circuits posed a challenge to the scientific community housed within corporate research organizations. In 1965, the same year as Gordon Moore's first "Moore's Law" publication, Jack Morton, Bell Labs' Vice President of Electronic Technology, summed up this tension. Morton warned of a coming "plateau" in electronics as the industry's technological progress outran the frontiers of science and urged the industry to make "repeated trips back to the 'fountain of youth' of basic science" to avoid the type of obsolecense that had struck other major industries (Morton 1965). In the decades following, despite the continued dominance of silicon integrated circuits, researchers in the industry's central research labs did try to couple the promise of new scientific breakthroughs with concerns over the "indefinite" extendibility of silicon into programs investigating new computing paradigms.[52] As documented by Mody (2016), IBM housed research programs in Josephson Junction computing and molecular electronics that both followed this blueprint. IBM's Robert Keyes published especially influential works that raised questions about the ultimate extendibility of silicon microelectronics (Keyes 1969; Keyes 1975). Researchers at IBM coupled these concerns with the promise of new technologies to build institutional support for their visions of alternative computing paradigms (Choi & Mody 2009; Mody 2016).

[52] Instead, the technical history is replete with examples of highly touted breakthroughs that made nary a dent on the commercial market. The tunnel diode, invented by Leo Esaki of Sony Corporation in 1957, was held up as a possible replacement to the transistor itself. An entire panel at the 1960 IEEE Solid-State Circuits Conference was devoted to the possibility of logic circuits built around tunnel diodes, but a market for the devices never materialized.

The 1980s marked a turning point for semiconductor research and development in the United States. With the growing challenge of Japanese competition, the first-wave of vertical disintegration, and the decreasing size and relevance of military R&D funding to commercial technology, industry embraced collaborative research and reduced funding of long-range research in-house: Firms funded both horizontal (SRC) and vertical (SEMATECH) collaborative research programs. At the same time, firms scaled down the scope and size of their central research labs (Macher et al. 1998; Macher et al. 1999). Long-range research on new device concepts was kept alive by mostly military funding (Choi & Mody 2009). For example, although envisioned as a program to speed up the development of military-specific ICs, the launch of the Very High Speed Integrated Circuits (VHISIC) program in 1979 precipitated a substantial increase in DoD funding for non-silicon integrated-circuit technology (ICE 1982). Similarly, the Office of Naval Research established the Ultra Submicron Electronics Research (USER) program in 1978 with the goal of "an electronics technology based on devices with 20 angstrom feature sizes" (Cooper 2007). Finally, Forest Carter took the leadership mantle for "molecular electronics" from his perch at the Naval Research Laboratory during this period (Choi & Mody 2009).

During the 1980s, researchers at the biggest industrial research labs in the industry (AT&T, IBM, and TI) actively used arguments about limits to scaling to urge investment and research in alternative devices such as low temperature operation (Dennard & Wordeman 1985; Dennard 1983) or devices based on quantum effects (Heilmeier 1984; Bate 1986; Capasso et al. 1989; Bate 1990). Texas Instrument's Central Research first began investigating quantum effect devices and architectures in 1982, under a program titled "Ultrasmall Electronics." Bob Bate, the program's manager, harkened back to Morton's 1965 piece on the need to "go back to the well"

of science to invent future electronics technologies (Bate 1990). An analysis of scaling trends by Texas Instruments researchers, supported by DARPA funding, concluded that "the advantages of geometry scaling do not continue to be attractive in the <0.5 um region" (Chatterjee et al. 1983). Subsequent publications by TI researchers in the nanoelectronics program specifically referenced this report's conclusions as its raison d'etre,

> "The rationale included the fact that continued downscaling would lead to size scales where quantum mechanics would effect [sic] electron transport. **Instead of attempting to design devices that operated in spite of quantum size effects, it is proposed that devices could be built whose operation was based upon these effects**" (Randall et al. 1989) [emphasis added].

Funding for TI's work in quantum effect devices came at least partially via the Department of Defense, initially through the Office of Naval Research and Army Research Office. Contract data from DARPA shows that TI received funding for six different projects through the 1990s totalling over $10 million.. By 1989 TI researchers began using the term "nanoelectronics" to describe the research program (Randall et al. 1989).

Government funding agencies put together the first broadly focused "nanoelectronics" programs in the early 1990s. DARPA's ULTRA Electronics program consisted of two phases, one running from 1991 through 1995 and the second phase ending in 1999-2000. The program supported a range of activities including compound semiconductors, novel devices (single-electron transistors and TI's work on resonant tunnel devices), new lithographic approaches (e.g. imprint lithography), and in later years projects related to molecular electronics. The program's primary recipients of funding were research labs at defense contractors and academics. Importantly, in reports surveying global nanotechnology programs put together by the interagency working group that would later go on to spearhead the National Nanotechnology Initiative, the DARPA ULTRA program's goals were described as producing "systems operating

at room temperature at speeds 10 to 100 times faster than current systems, denser by a factor of 5 to 100, and lower-power by a factor of more than 50" (WTEC 1998). Thus, DARPA ULTRA's goal was to maintain Moore's Law with a successor to the CMOS silicon integrated circuit.

Similarly, a European research program started in 1996, The Advanced Research Initiative in Microelectronics (MEL-ARI), pursued a very similar research program to that of DARPA's ULTRA electronics. The "nanoscale integrated circuits" program funded projects pursuing single electron memories, molecular electronics, silicon quantum electronics (resonant tunnel diodes), and imprint and STM lithography approaches. The two programs were aware of each other; Gernot Pomrenke, ULTRA's program manager during its second phase, attended the MEL-ARI conference in 1997.

During the same decade that government funding agencies were constructing a community of nanoelectronics researchers and approaches, the industry's collaborative organizations were focusing its associated research community on the accelerating pace of Moore's Law in silicon integrated circuits. By the middle of the 1990s, the primary challenge facing the U.S. semiconductor industry was no longer international competition, but the difficulty of maintaining the industry's business model and its associated technological trajectory. The U.S. semiconductor industry entered this period with more stakeholders across a vertically disintegrating industry that pursued less and less in-house fundamental R&D. Startups had embraced the separation of design from manufacturing, and integrated device manufacturers had drastically cut the basic research budgets of their central research labs and focused their research efforts toward applied research in support of their existing businesses (Macher et al. 1998). Throughout the 1990s, the industry engaged in a collaborative technology direction-setting process, which, in turn, shaped academic research and industrial commercialization

strategies (Schaller 2004). By the end of the decade, the same institutions established by the industry to combat international competition and cooperate on advancing the CMOS paradigm, including the Semiconductor Industry Association, Semiconductor Research Corporation, SEMATECH, and the National Technology Roadmap for Semiconductors, shifted the industry's focus toward increasingly urgent long-term technology research needs of replacing CMOS altogether.

Prior to 1994, the industry's roadmapping exercises focused on 10-year extrapolations of the industry's existing technology trajectory. In what became known as "Moore's Law," industry pioneer Gordon Moore showed in 1965 that the number of components on a commercial integrated circuit had doubled roughly every 18 months since ICs were introduced to the market and argued that there were no reasons not to expect this trend to continue indefinitely (Moore 1965). Before long, maintaining Moore's Law became both the technological and economic driver of the industry[53] (Brock & Moore 2006). For MOSFET architecture, which became the industry's dominant design beginning in the 1980s, the shrinking of feature sizes resulted in devices that were faster and more energy-efficient and conformed to Moore's Law (Dennard et al. 1974).

The industry's 1994 National Technology Roadmap for Semiconductors briefly identified three areas for long-term research: "nano-metrics metrology, nano-fabrication techniques, and new (post shrink) device structures" (NTRS 1994). The report's glossary defined "post-shrink" as "a descriptor of performance-enhancing methods for integrated circuits that will become

[53] The late eminent historian of technology, Thomas P. Hughes, coined the expression "technological momentum" to describe such a phenomenon (Hughes, 1971). The sociologist of technology Donald MacKenzie specifically described Moore's Law as a "self-fulfilling prophecy" for the semiconductor industry owing to the degree to which it drove the industry's technological trajectory and the capital markets' mode of gauging the performance of firms such as Intel and AMD (MacKenzie, 1996).

applicable when the physical limits of scaling (shrinking the dimensions of the semiconductor structure) are reached" (NTRS 1994). The roadmap outlined two parallel paths forward for devices:

> "The first is directed toward evolutionary devices fabricated using Si technology. The second explores revolutionary devices, possibly using alternative devices and materials to discover new principles of operation that may potentially be retrofitted into CMOS" (NTRS 1994, p.62).

Evolutionary concepts included new geometries for the CMOS structure (e.g. dual gate and vertical MOSFETs) while revolutionary devices were the "quantum device concepts" that had been pursued by the industry's central research labs in the 1980s and were now finding support from DARPA. At least one of the contributors to the roadmap's Process Integration, Devices, and Structures chapter was Stanford's Jim Plummer. Plummer happened to also have been Principle Investigator on a grant in DARPA's ULTRA program.

The 1994 report set the scene for the coming evolution and potential revolution in microelectronics but highlighted the challenging long-term funding environment, noting especially that the decline of industrial research laboratories "[left] a major gap in the U.S. [research] infrastructure" (NTRS 1994). An analysis commissioned by SIA and SRC following the release of the 1994 NTRS found industry and government funding for long-term research insufficient to address the technology obstacles identified by the NTRS and argued that universities could address a considerable portion of the "research funding gap" (Bodway et al. 1995). In late 1994, after SEMATECH's board voted to forego additional federal funding, the Semiconductor Technology Council was established with representatives from government (NSF, Departments of Defense, Energy and Commerce), industry (semiconductor manufacturers, suppliers and users), and academia "to help guide research and development in semiconductors"

as well as "link assessment by the semiconductor industry and national security needs for cooperative investments… and align industry and government contributions for new semiconductor research and development efforts" (HR2401 1994).

In 1996, the Semiconductor Technology Council proposed a new program for funding university research aimed at closing the research funding gap identified several years earlier by the NTRS. Semiconductor manufacturers, equipment suppliers, and the federal government (initially through DUSD[54] and then DARPA) jointly funded a new SRC subsidiary in 1997, the Microelectronics Advanced Research Corporation (MARCO) to oversee the Focus Center Research Program (FCRP), to fund research centers to "seek creative options for the solution of key technology challenges so that the industry can keep pace with the cadence of Moore's Law" (SRC 1998). The FCRP was also a response to growing frustration amongst SRC-associated academic researchers at the lack of academic freedom available via SRC grants. The FCRP funded university centers where center leadership had wider latitude to determine research directions. For much of the following decade, however, FCRP's research funding remained within the Si-CMOS paradigm.

The 1997 edition of the National Technology Roadmap for Semiconductors once again highlighted the challenge of adequately funding long-term research but did not directly comment on the need for and challenge of "post-shrink" technologies. The report's primary foci were the challenging economics of continued scaling and impending challenges of integrating new materials throughout the CMOS architecture (e.g. new gate dielectric, gate electrode and interconnect metals). For technology node projections beyond 2006 the report simply called for "the most innovative research, including serious consideration of paradigm changes" with no

[54] Defense Undersecretary of Defense

further specificity (NTRS 1997). In the decade following the 1997 report, the semiconductor industry devoted considerable engineering effort to overhauling the materials employed in CMOS architecture in an effort to continue scaling. IBM introduced copper interconnects in 1997. Intel brought strained silicon source and drain to market in 2003, and high-k dielectric and metal gate in 2007. These new materials were engineered to mitigate complications from quantum effects and materials limitations at increasingly smaller dimensions. Throughout the 1990s and early 2000s, the industry's response to the presumptive anomaly identified in the 1970s and 1980s was not a radical new technology direction but rather a series of modular changes within the existing CMOS architecture (Henderson & Clark 1990).

By the turn of the century, at the launch of the National Nanotechnology Initiative (NNI), the nascent nanoelectronics community largely existed outside the purview of the mainstream semiconductor industry. DARPA initiated several programs in alternative device technologies that built on the results of the agency's ULTRA program. These programs included spin-based electronics (SPINS), molecular electronics (Moletronics, MoleApps), and steep sub-threshold devices (STEEP). Europe's funding agency also created a follow-on program in 1999-2000, Nanotechnology Information Devices, that focused primarily on self-assembly, molecular electronics, and DNA-computing techniques (Compano, 2001).[55]

Meanwhile, the SRC's primary research program, mirroring the focus of the semiconductor industry roadmap, remained oriented toward addressing the industry's needs with respect to continued scaling of CMOS. Arrayed into a group of "science areas," SRC's main research program paralleled the organization of the industry roadmap while also appealing to a wider group of firms in the increasingly specialized industry (Burger 2000). Consequently,

[55] The European research program published a "Nanoelectronics Roadmap" in 1999 and 2000.

SRC's initial solicitations for "beyond CMOS" concepts were essentially ad-hoc requests made to research centers long-affiliated with SRC. The Cornell Nanoscience Center's (CNC) summary report to SRC for 1999 describes how SRC leadership solicited a CNC proposal for "CMOS replacement" devices during the 1999 timeframe:

> "The focus during the first half of the reporting year was on 20 nm CMOS, but it shifted in the second half towards the search of a CMOS replacement. The mandate to explore a CMOS replacement was given by the SRC in a meeting at Cornell during the spring of 1999. At the same time the new roadmap, the 1999 ITRS, was being finalized by the semiconductor industry and SRC underwent a science area reorganization. All of this produced uncertainty and turmoil among faculty and students at Cornell. Subsequently, Cornell developed a new unique far-reaching theme, SET/CMOS, and prepared an integrated proposal with the title 'Silicon Based Nanoelectronics in CMOS Environment.' This proposal was submitted to the SRC in September 1999 and presented [sic] orally in the October 1999 annual review. The proposal was criticized at the same time for not reaching far enough and for not providing desired CMOS replacement characteristics. Consequently the proposal was not funded" (Krusius 2000).

Despite the CNC's setback, SRC's research program in beyond CMOS technologies began in earnest with the establishment of two small programs in 1999 and 2000, Cross-Disciplinary Research and the Advanced Devices Thrust (ADT) both of which funded a variety of different exploratory concepts. The conceptual basis for the programs was laid out in a 2000 report published by the SRC, "Research Needs in Basic Science of Semiconductors." The report's authors argued – optimistically – that "the physical and chemical understanding necessary to reach the ultimate limits of CMOS technology also will provide the basis for inventing new technologies that will eventually supplement CMOS" (Cavin, Herr, Zhirnov 2000). This dual mandate is mirrored by SRC's funding for "ultimate CMOS" and beyond CMOS technologies around this time. SRC's James Hutchby managed the ADT program with industrial liaison support from Intel's George Bouriaonoff. The modestly funded program focused on two parallel tracks: extending CMOS through non-classical structures and

investigating novel device concepts "that have the potential to process, store and communicate information at a speed, density and energy efficiency that greatly exceeds (100X) those projected for silicon CMOS at the 22-nm node" (SRC 2000). The ADT program's goals more or less echoed DARPA's ULTRA program from the 1990s.

SRC's research solicitation offered few specifics about what was considered a novel device. Instead, a footnote simply offered a definition:

> "However, 'novel device' more likely would be associated with entirely new concepts for and technologies related to information processing, storage, transmission (i.e., computation, storage and interconnection). Examples of currently identified structures that might be considered as future CMOS replacements include (but are not limited to) single electron transistors and coulomb blockade based devices; quantum dots, wires and arrays. Novel devices most likely will require new and novel circuit architectures and new materials and fabrication techniques." (SRC 2000)

Thus, SRC's initial funding of novel device concepts largely shared the two-pronged approach put forth in the 1994 NTRS. Additionally, the existing technologies highlighted by the solicitation were outgrowths of DARPA's ULTRA electronics program, which itself was inspired by research pursued in the central research labs of several semiconductor firms in the 1980s. SRC contract data shows that several small grants were made under the ADT program in 1999 and 2000. Included in these grants were alumni researchers and topics from DARPA's ULTRA program, including molecular electronics by Mark Reed at Yale and single electron transistors by Konstantin Likharev at SUNY Stony Brook. Through the end of the century, the industry, acting through the SRC, followed the lead of government funding agencies in approaches to "beyond CMOS" technologies.

Beginning in the early 2000s, however, the industry's approach toward "beyond CMOS" shifted from surveying and monitoring external efforts toward cultivating ideas and managing

research directions as the industry responded to the emergence of a new presumptive anomaly and the research opportunity presented by the establishment of the NNI. The SRC organized an advisory board for the ADT program consisting entirely of industry representatives from member companies, one representative from NIST (David Blackburn) and one academic (MIT's Dimitri Antoniadis). ADT leadership – Blackburn, Antoniadis, Hutchby, and Bourianoff – organized a series of "nanotransistor" workshops that brought together government agencies (NIST, NSF, NASA) and a select group of academic and industry researchers to examine questions related to scaling limits for MOSFET transistors. SRC researchers also coordinated with the international community of researchers considering similar questions. In 1999, the SRC hosted the International Workshop on Future Information Processing Technologies (IWFIPT) jointly sponsored by the NSF, European Commission, and Japanese funding agencies. A report by the SRC's Ralph Cavin summarizing the 1999 IWFIPT program shows how the conference was motivated by the soon-to-be-published conclusions of the 1999 ITRS on looming hurdles to continued development of the CMOS platform and included presentations on quantum computing, single electron memories, and molecular electronics.

These activities culminated in the authoring and inclusion of a new chapter in the 2001 ITRS, "Emerging Research Devices." The structure of the chapter followed the same two-pronged approach followed by the SRC: "non-classical CMOS" and "Emerging Logic" devices. The prospective device technologies surveyed in the "Emerging Logic" section were primarily those receiving funding from the existing government nanoelectronics programs: resonant tunnelling diodes, single-electron transistors, quantum-cellular automata, and molecular electronics. The authors of the chapter noted that many of the "Emerging Logic" devices were "speculative" and not directly competitive with CMOS, an assertion supported by the chapter's

rough quantitative comparison of emerging device capabilities to CMOS (ITRS 2001). The quantitative comparisons lacked a single device that projected to be strictly superior to CMOS in every major metric (cost, speed, energy, and density). The majority of the American contingent to the original PIDS & ERD chapter in 2001 were members of the SRC ADT advisory group. European members were associated with the MEL-ARI project, including Ramon Campona, MEL-ARI's director. The table below summarizes the simultaneous membership of key individuals in several of these organizations. Individuals are demarcated as an X if they were listed as a presenter or contributor to the organization. Early consensus organizations brought together input from two groups. Those affiliated with SRC research programs (FCRP and ADT) and program managers and researchers from early government nanoelectronics programs (ULTRA, MEL-ARI). Data from Japanese contributors is limited.

Table 4 - Visual summary of key individuals that informed early nanoelectronics roadmaps and consensus documents in the years prior to the launch of the NRI.

		Gov Nano		SRC		Consensus Orgs		
Name	Organization	ULTRA	MEL-ARI	FCRP	ADT	IWFIPT	TNT	ITRS ERD
George Bourianoff	Intel/SRC				X	X	X	X
Bob Doering	TI				X	X		
Marc Van Rossum	IMEC		X			X		
Kristin De Meyer	IMEC		X			X		X
Ramon Campano	EU		X				X	X
Dan Radack	DARPA	X		X		X		
Gernot Pomrenke	DoD	X				X	X	
Jim Hutchby	SRC				X		X	X
Ralph Cavin	SRC			X	X	X		
Toshiro Hiramoto	Univ. Tokyo					X		X
Mike Forshaw	UCL	X	X					X
Kang Wang	UCLA			X	X	X	X	
Dimitri Antoniadis	MIT			X	X			
Year of Formation		Pre-1999	1997	1999	1999/2001	1999/2001	2002	2001/2003

SRC researchers reacted quickly to the relatively dour projections for the prospective successors to CMOS. At the 2001 IWFIPT, George Bouriaonoff of Intel presented a talk summarizing the Emerging Research Device chapter from the ITRS. The next year, Ralph Cavin of the SRC applied for and received a grant from the NSF for a session at the *Trends in*

Nanotechnology Conference in order to "visualize the direction needed to approach the long-term goals of developing nanoelectronic technologies." SRC researchers presented a talk, "***Nanoelectronics -- Current Status and A Point of View***" that laid out the SRC's view on existing "beyond CMOS" device concepts and the need for "unconstrained research" into solutions. Their presentation walked through the SRC's developing thesis on the need for a new paradigm of research in "beyond-CMOS."

Significantly, the fully revised ITRS, issued in 2003, placed the ERD chapter as its principal focus. The report's executive summary highlighted the ERD chapter's discussions of post-CMOS devices as "pav[ing] the way to a complete technological revolution looming ahead towards the end of the next decade" (ITRS 2003). The chapter went beyond a broad survey of emerging device concepts to include "a balanced, critical assessment of these emerging new device technologies" that "provides an industry perspective" (ITRS 2003). The 2003 ERD chapter's final section, "Emerging Technology – A Critical Review," was built around a fundamental analysis of binary logic switching conducted by Victor Zhirnov, Ralph Cavin, James Hutchby, and George Bourianoff, all key staff members of the SRC and all core members of the ITRS ERD working group. This group had presented the same thesis at the Trends in Nanotechnology Conference the year prior. In an ideal case of presumptive anomaly, as defined by Constant (1980) (e.g. an understanding of future conditions under which the existing silicon CMOS system would fail), their analysis concluded, "even if entirely different electron transport devices are invented for digital logic, their scaling for density and performance may not go much beyond the ultimate limits obtainable with CMOS technology, due primarily to limits on heat removal capacity." The chapter called for additional research on "alternate physical mechanisms for device operation" (Zhirnov et al. 2003). Building on the conclusion of the above analysis, the

ERD report concluded that none of the existing research devices were "viable emerging logic technologies for integration" (ITRS 2003).

The SRC and industrial leadership responded to the lack of viable alternatives by engaging in a collaborative direction-setting process to identify new avenues for research. SRC's "Novel Device Task Force" issued a "Research Needs for Novel Devices" document in May 2003 that summarized the situation:

> "During the last several years there has been a huge amount of government-sponsored research loosely described as nanotechnology. This effort has lacked organization and direction and consequently is of limited immediate value to the semiconductor industry. However, the associated body of knowledge provides a rich resource that can be used as a starting point for targeted research programs" (SRC 2003)

In the months following, semiconductor industry leadership, acting through the SIA and the SRC, began to coordinate a response to the conclusions of the SRC's Novel Device Task Force document. In July 2003, the SIA convened a subcommittee of senior industry executives from TI, IBM, Intel, Motorola, Micron, and SRC to organize a "nanotechnology strategy." The goal was to leverage vast new public funds for nanotechnology research allocated by the 2001 National Nanotechnology Initiative. Beginning in October 2003, SRC and NSF organized three industry-academia-government workshops, Silicon Nanoelectronics and Beyond (SNB), to discuss research directions for addressing the industry's looming challenges. The workshops, organized by Intel's Paolo Gargini and the NNI's Michail Roco, laid the groundwork for the industry's increasing influence over all beyond CMOS research activities in the US. Participants in the workshops established five joint NNI-SRC working groups, the "International Conferences on Communication and Cooperation" and, added in 2006, the "International Planning Working Group for Nanoelectronics" to map and coordinate global research efforts. Members of these groups played influential roles in shaping both national nanoelectronics programs and also the

official position of the semiconductor industry vis a vis the ITRS Emerging Research Device chapters. The SRC also signed a Memorandum of Understanding (MOU) with the NSF to review any proposals made to the NSF's "Silicon Nanoelectronics and Beyond" program beginning in 2004. Under the MOU, relevant proposals to the NSF were reviewed by SRC chosen industry researchers. In addition to inviting these researchers to "voluntarily participate" in the SRC's technology transfer processes, some of the projects received additional funding from the SRC (NSF 2004).

3.2. Shaping a Response to Moore's Wall

These two efforts, the SNB Workshops and the SIA Technology Committee's nanotechnology strategy discussions, jointly shaped the formation of the industry's response to the expected future under which the existing silicon CMOS system would fail as laid out by the ITRS ERD group, and encapsulated in the Zhirnov et al publication. While the SNB workshops helped to define the beyond CMOS research agenda, the SIA Technology Strategy Committee laid the groundwork for a new organization, the Nanoelectronics Research Initiative, founded specifically to address the challenge of identifying computing devices that did not rely on electron charge.

The SIA Technology Committee, in response to a survey request from PCAST about NNI objectives, considered a range of organizational forms to help the semiconductor industry tap government investments in NNI. The different organizational forms varied primarily along the relative shares of industry and government funding. In January 2004 the SIA Technology Strategy Committee considered a proposal put forth by Intel's Paolo Gargini for an "Industrial Research Institute" to be funded primarily through the NNI, i.e. with public funding, in the range of $600 million per year. It would be staffed by industry researchers and visiting academics

drawn primarily from the existing FCRP community (Gargini 2004). In June 2004, however, the SIA board instead approved a new program, the Nanoelectronics Research Initiative. The approved NRI structure was conceptually much closer to SRC's existing FCRP multi-university center model than the "Industrial Research Institute" discussed at the January meeting. SIA documents indicate the initial approval came without having secured any government funding but aimed for a 90-10 government-industry funding ratio, a departure from previous SRC programs, which were majority-funded by industry (SIA 2004). Per the SIA Board, the NRI's research objective was to "discover and reduce to practice via technology transfer to industry novel non-CMOS devices" by 2020 (SIA 2004). NRI organizers described the program as "switch centric" research with the expectation of expanding the set of potential options available to industry and defining criteria for evaluating different options (Bourianoff 2004).

The NRI's research program directly paralleled the concluding recommendations of the SNB workshops. According to Intel's Bourianoff the SNB workshops were "not able to reach a clear consensus on which alternate state variable was most promising" and instead "define[d] 13 research vectors participants believed were essential to make progress in the search for the next switch beyond CMOS" (Bourianoff 2006). NRI planners adopted the first five[56] of the thirteen research vectors identified by the SNB workshops as the core of its program. In a white paper introducing the NRI, Intel's George Bourianoff and IBM's Tom Theis explicitly referenced the implications of Zhirnov et al.'s (2003) model as the basis for the choice of these research vectors. Furthermore, they argued that the choice of research vectors would also directly shape the possible design space for NRI researchers: "the search for the next logic switch beyond CMOS

[56] (1) Computational State Vectors, other than Electronic Charge, (2) Non-equilibrium Systems, (3) Novel, Non-charge Data Transfer Mechanisms, (4) Nanoscale Phonon Engineering for Thermal Management, (5) Directed Self-assembly of Such Structures.

will need to embrace some radical paradigm shifts and focus on key elements that limit conventional micro electronics" (e.g. limits to heat dissipation) (Bourianoff & Theis 2004).

Concurrent with the efforts to structure the orgniazation and define a research agenda, industry researchers associated with the NRI conducted a survey of existing NSF funded Materials Research Science and Engineering Centers (MRSEC) and Nanoscale Science and Engineering Centers (NSEC). To assess how well existing research center research programs and capabilities aligned with the proposed NRI research agenda. Bourianoff and Theis summarized the industry's view of the relationship between the research vectors and the existing university-run scientific research programs funded primarily by the NSF:

> "While this [the NRI's] research agenda seems quite radical when stated as a coherent whole, in fact many elements of it already exist in the research community. Taken together, they provide the next level of description for what the 'next logic switch beyond CMOS' might look like. The discovery model of research nurtured by the NSF seems to have laid the foundations for this new technology in an uncorrelated, bottoms up, curiosity-driven fashion and *our task now is to synthesize selected elements into a coherent research plan.* These selected elements would form the 'goal' of a research program planning process. After performing an assessment of the existing research inventory in university NSEC centers, MRSEC centers, MARCO centers and other centers of excellence, it will be possible to do a "gap analysis" of the requirements to reach the goal and create a plan to execute the gap analysis." (Bourianoff & Theis 2004) [emphasis original]

The concept of a gap analysis had been used previously by the industry in creating the SRC's FCRP model following the conclusions of the 1994 NTRS. While the FCRP model grew out of identifying gaps between industry and academic research, this NRI research planning analysis focused on research funded by the NSF and existing SRC programs. However, committee planning documents highlight the industry's delicate balancing act with respect to existing government support for beyond CMOS research from NSF and DARPA. The authors of these documents stress that the proposed initiative should not be "interpreted as negative

criticism" of and should "synchronize" with existing public investments. In follow-on presentations NRI leaders presented the NRI as building on the results of federal investments in science, as shown below in figure 12.

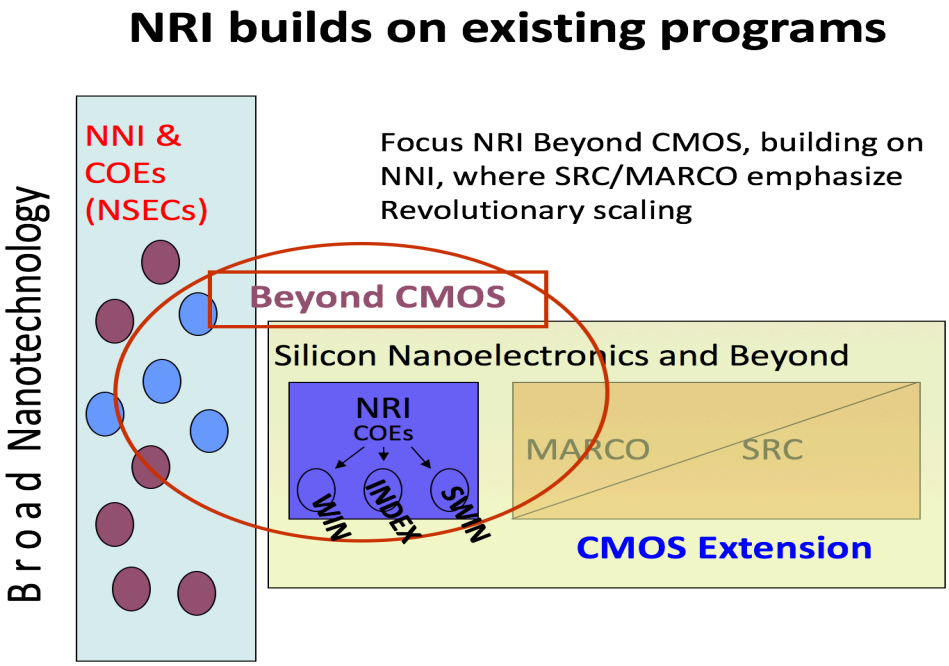

Figure 12 - Description of NRI program relative to existing government programs. Source: (Bourianoff et al. 2006).

Not mentioned explicitly in Bourianoff and Theis's white paper was DARPA-funded work in spintronics, molecular electronics, and other concepts dating back to the ULTRA programs of the 1990s. But one NRI planning document describes the DARPA perspective as, "What are the new credible device concepts beyond CMOS that are not already being funded?" (SIA, 2004). At the same time, SRC researchers were explicit in saying that the existing research streams – many of them dating back to the DARPA ULTRA program – were unlikely to yield viable alternatives,

> "Few of the proposed new electron transport devices (RTDs, SETs, RTTs, molecular devices, etc.) emerging from research appear to satisfy all of the criteria listed above for

> integration on the silicon CMOS platform, at least in their current configurations" (Cavin & Zhirnov 2004).

Although established in 2005, the NRI did not begin funding research until 2006 with the founding of its first multi-university centers. The formation of these centers highlighted two distinct themes: the need for industry to prove its sincerity in funding long-term non-silicon-CMOS research and the industry's reliance on university researchers with whom it had long-standing relationships. Interviews with center directors and NRI leadership indicate that firms approached the professors, encouraging them to organize a program around NRI's goals. However, several industry researchers remarked that academic researchers initially responded with skepticism about NRI's mission and doubted the sincerity of the industry's quest for an entirely new device concept. Founding NRI director Jeff Welser summed up the disconnect between academic leaders and industry members:

> "For the first couple years, our problem with them was they would not think far enough out. They were so used to being basically beat up by us saying, . . . [when] they'd come up with some idea . . ., 'That's ridiculous, that will never work.' Because that was the SRC program and the industry mindset had been very much on, 'I need something for the next N plus two nodes or whatever, so don't tell me about this weird material.' So, they didn't believe us. Even though we said, 'We want something that's not a FET,' everything they showed us was an FET. It was different materials, but everything was an FET. And, we were like, 'But we said we don't want an FET.' And they said, 'Yeah, yeah, but we don't believe you.'" (Welser 2013)

The NRI did not receive federal funding during the program's first year. Instead, in each case a member firm formed a joint-proposal with universities located in the home-states of the firm and that state's government. A failed proposal to the NSF for a spintronics center led by a team from UCLA in coordination with Intel became the core of the first NRI center, the Western Institute for Nanoelectronics (WIN). TI and IBM worked closely with universities and state governments in Texas and New York, respectively, to emulate the model put forth by Intel and the WIN center. In addition to funding from the consortium, each center received funding from

state government, as well as start-up funding and equipment donations from the sponsoring firm. In California and Texas, established state-level programs (the UC Discovery Program and the Texas Emerging Technology Fund) were the source of funding for the NRI centers. IBM played an instrumental role in helping the MIND center get support from local and state governments.[57]

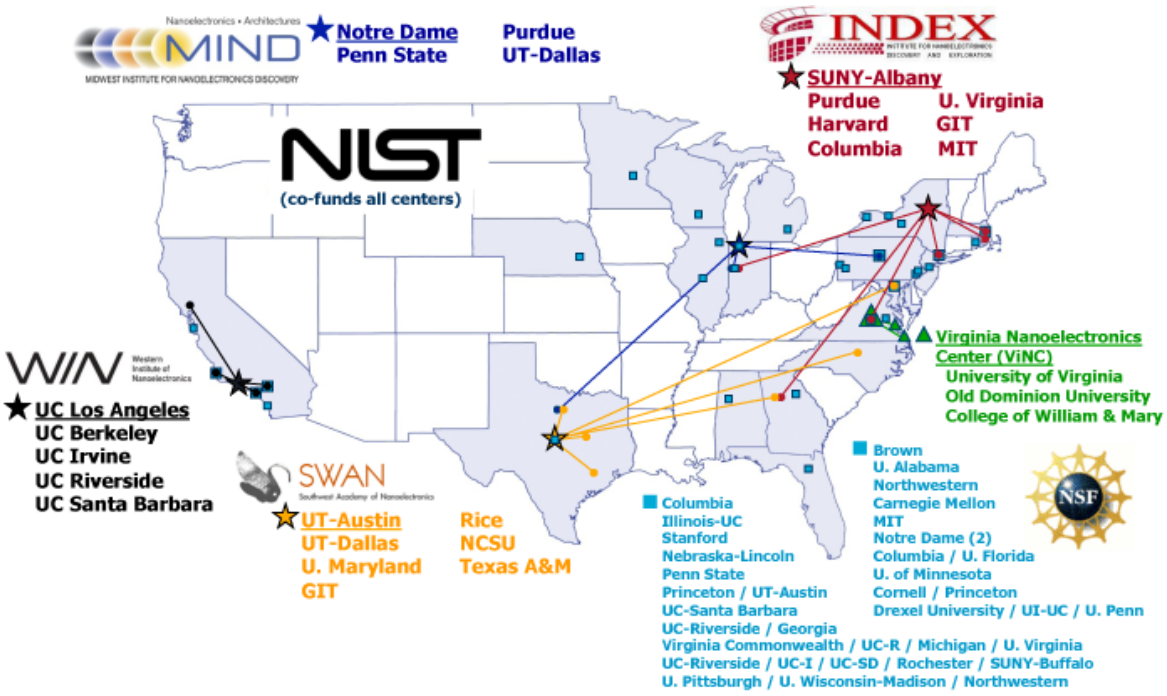

Figure 13 - Nanoelectronics Research Initiative project and center map as of December, 2012. Source: (SRC 2012).

While the NRI's member firms helped drive the fundraising process by leveraging their local clout, the research agenda was driven by center directors chosen by the organization's leadership. NRI center directors were chosen to be "typical engineering professors" who had a long history of interaction with industry research either through SRC or in industrial labs (Welser 2013). Kang Wang and Sanjay Banerjee, directors of the WIN and SWAN centers, respectively,

[57] The stated rationale behind state-level funding for the centers was job-creation and technology commercialization. Concerns regarding job-creation were especially acute in the wake of the economic downturn. Notably, given the long-term nature of the R&D and the location of the member firms, it is unlikely that the technologies would actually benefit the local economy beyond the university professors funded and the associated technicians.

had each received more than 25 contracts from SRC prior to 2006. Wang was also the director of an FCRP center, FENA and a contributor to the SRC's "nanotransistor" workshops in the late 1990s. Banerjee had worked at Texas Instruments in the 1980s before entering academia. Founding director of the INDEX center, Alain Kaloyeros, had led the development of the Albany Nanotech Center and worked closely with IBM and other semiconductor firms since the mid-1990s. Alan Seabaugh, the director of the MIND center at Notre Dame, had worked at Texas Instruments and Raytheon Central Laboratories in the nanoelectronics group before moving to Notre Dame in 1999. The plot below visualizes the extent of NRI researchers relationship with the SRC prior to their involvement in the NRI.

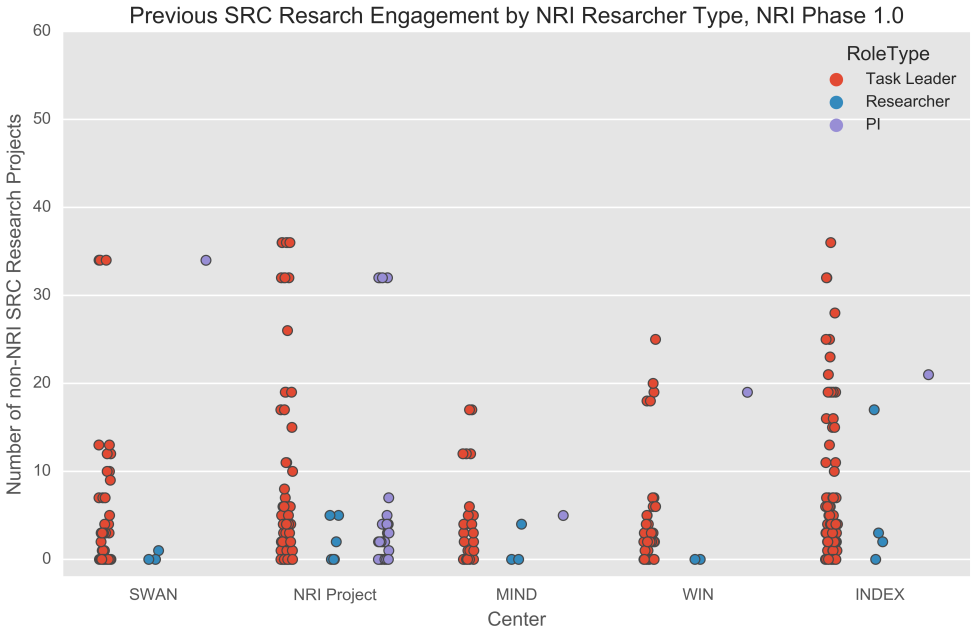

Figure 14 - Comparing NRI researchers by center and type based on their history with the SRC. Many researchers involved with the NRI had never worked with the SRC previously. Source: SRC task data.

Importantly, their decades of experience with industry research meant that center directors understood the industry's looming challenges. In the case of the SWAN center,

leadership organized workshops on CMOS device operation to bring scientists unfamiliar with electrical engineering concepts up to speed on the challenge facing the industry. These seemingly simple interactions served to educate scientists on basic boundary conditions for identifying a successor device and contributed to the evolution of researchers' device proposals (Welser 2013).

Federal funding for the NRI centers became available in 2007 through NIST. In 2007 NIST issued a call for funding a "Consortia for Post-CMOS Nanoelectronics Research Program" tailor-made for the NRI (NIST 2007). NIST funding through 2012 was $2.75 million a year, and NRI member firms contributed annual dues of around $1 million each to fund projects at both the centers and through the NSF partnerships. In 2008, the NRI established a fourth multi-university center headquartered at Notre Dame. The MIND center, the winner of an open solicitation, did not receive any equipment donations from member firms; instead, the center received considerable funding from the university, state, and even city. (See figure 15 for funding amounts for the four centers and figure 13 for a map of NRI funded centers and projects).

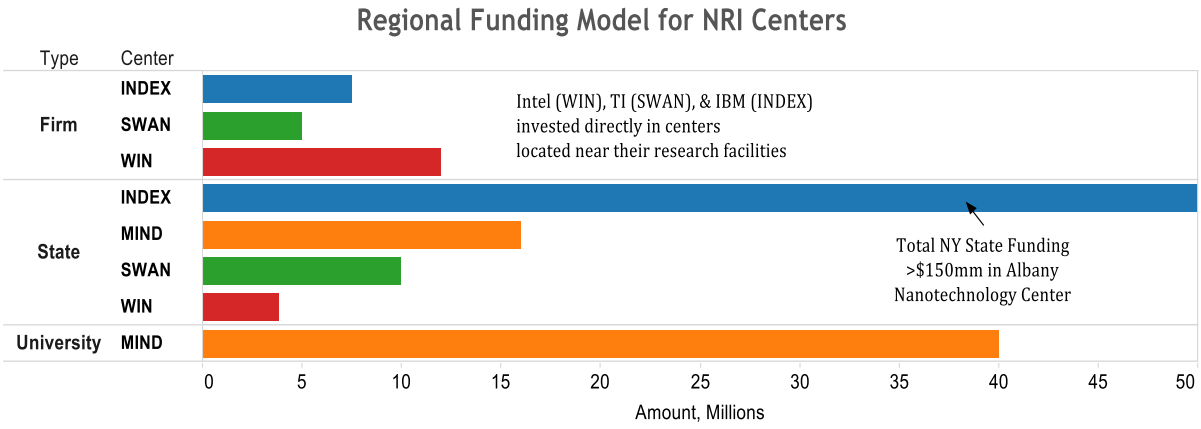

Figure 15 - NRI centers featured co-investment from firms and state governments. Source: (SRC 2005; SRC 2006b; SRC 2008)

In addition to the multi-university centers, NRI also began co-funding joint awards at NSF MRSECs and NSECs[58] beginning in 2006. Unlike NRI center-based research, which focused teams of researchers on specific approaches to post-CMOS devices, NSF projects were more diverse. Figure 5 below summarizes how NRI leadership saw the differences in focus between existing MRSEC and NSEC programs, NRI Centers, and NRI-NSF projects relative to their focus on the research vectors identified by the SNB workshops. The NRI-NSF projects received 90% of their funding from the NSF. Unlike NRI centers which had dedicated assigness or liaisons, contact with NSF projects primarily occurred only during annual industry reviews. IBM's Jeff Welser, director of NRI from 2006 through 2012, noted that NSF project research aimed to expand the scope of the NRI program.[59] As Tom Theis, an IBM research scientist who had long been involved with SRC and who succeeded Welser, explained:

> "Well, even IBM with the resources we have here can't consider putting enough people on this problem. ⋯ Because of NRI, you've got 34 universities with top people involved thinking about this problem. IBM can't match that . . . And so, that [i.e., the larger university research] community is going to come up with ideas that IBM is not going to come up with. Now, IBM researchers can inject their own ideas into that mix. That's happening⋯, but for the amount of money that IBM puts into NRI, you could not do anywhere close to that amount of research within IBM" (Theis 2013).

Similar to previous SRC programs, the NRI leveraged industry researchers for evaluating and providing feedback on ongoing research projects. Industrial researchers engaged with NRI academic research through a variety of mechanisms. Two governing boards, the Technical Program Group and Governing Council, were comprised of researchers and research managers from industry and government. They formed the core group of technical reviewers who

[58] Materials Research Science and Engineering Centers & Nanoscale Science and Engineering Centers

[59] NRI's newest regional center, CNFD, established in 2013 and located at the University of Nebraska, first began as an NSF co-funded project.

evaluated the merits of different proposals as well as offering input on management issues throughout the duration of the program. Several members of the NRI's TPG were also instrumental in defining its initial research agenda. Industry researchers also engaged directly with academic research as both industry liaisons and assignees. NRI assignees were industry researchers designated by member firms to work on-site at academic research centers in coordination with NRI researchers.

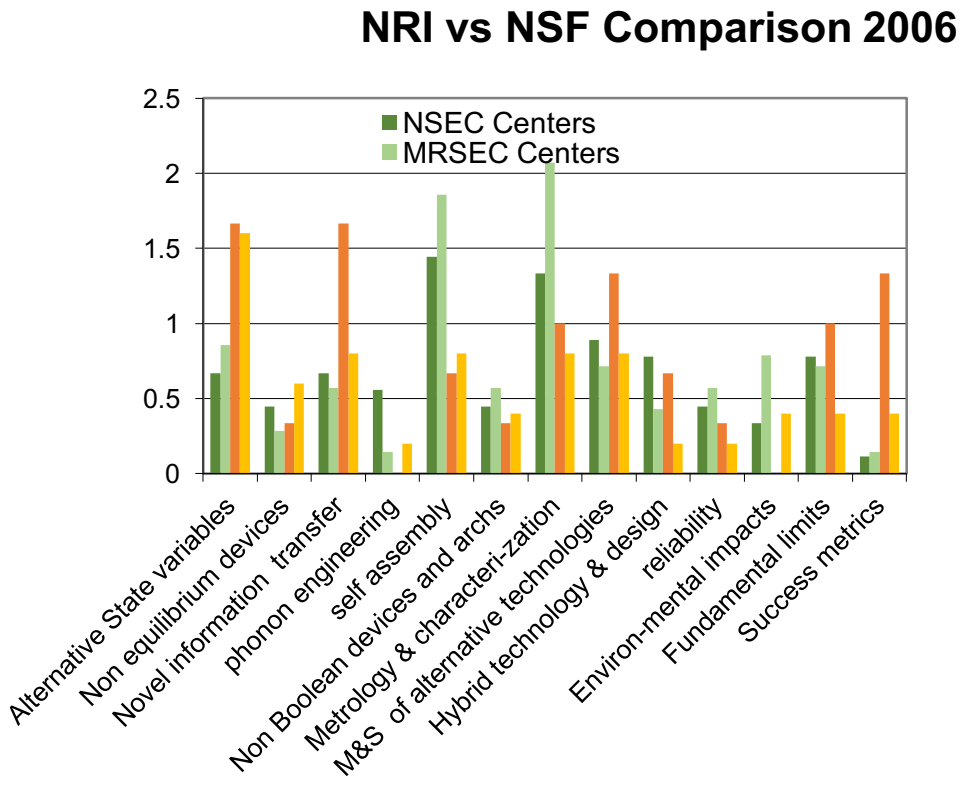

Figure 16 - Summary of SIA's NSEC and MRSEC visit reports. This analysis by NRI organizers shows the differences in the NRI program's objectives relative to bottoms-up centers already in existence. Compiled by author using data from Bourianoff, 2006.

In interviews, academic researchers noted that while joint work was published, the industry researchers functioned primarily as employees of their firms. One Intel assignee worked with the WIN center at UCLA for 5 years, where he described his role as "getting people to collaborate and integrating all spin-related research projects" (Assignee 2013). TI's assignee

worked primarily with the SWAN center in Texas and co-patented six inventions with SWAN researchers, nearly half of the NRI's total patent output. Both assignees noted that they limited their own patenting in order to manage relationships with academic researchers and stressed that they were most effective as mentors for projects and students. The ongoing feedback industry researchers provided to participants about research directions also influenced research trajectories. Several participants described the assignee role of giving feedback as "providing focus" for academic researchers on directions most likely to be fruitful. Some academic researchers occasionally bristled at the loss of research freedom, and this was a source of tension early on in the program. In interviews, academic researchers who had accepted NRI funds were coy about discussing the demands placed on them by NRI's funding.

The program structure defined by 2008, with the infusion of funding from NIST, would last for another three years. In 2011, NRI leadership asked the four centers to refocus their programs on a smaller number of device concepts. This was described as phase 1.5 of the NRI program and led to a winnowing of projects and device approaches. For example, the INDEX center based at SUNY Albany went from supporting work by 21 Principal Investigators at eleven universities on four device concepts to supporting 14 Principal Investigators working on two device concepts at seven universities. Figure 17 below shows the evolution of total funded projects under the NRI during this period.

The SRC re-competed both the NRI and FCRP programs in 2012 with an open proposal process. The newly recompeted FCRP program was renamed STARNet and continued to draw funding from integrated device manufacturers, equipment suppliers, defense contractors and DARPA. Several of the projects under research at the MIND and WIN centers during NRI's phase 1.0 and 1.5 were moved into STARNET centers. NRI phase 2.0 continued to receive

funding through NIST but was reduced to three centers. Both the SWAN and INDEX centers were renewed and a third program, the Center for NanoFerroic Devices (CNFD) headquartered at the University of Nebraska, Lincon, was established. Research leaders at CNFD first joined the NRI as an NRI-NSF project in 2008.

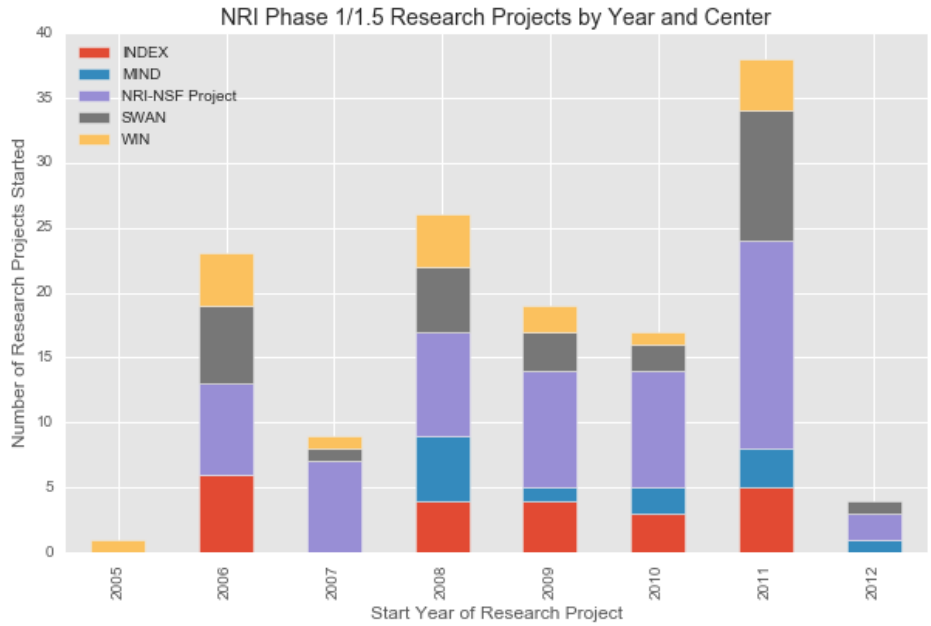

Figure 17 - Total projects funded per year by center at the NRI, 2005 - 2012. Source: NRI task data, compiled by author.

3.3. Sizing Up the NRI

The landscape of research activities in the field of emerging computing technologies is significantly broader than what has been funded by the NRI. Researchers involved in the NRI received funding support from multiple agencies simultaneously and there continues to be significant market, scientific, and technological uncertainty about the industry's near and long-term future. It remains today unknown whether any of the technologies investigated by NRI researchers will reach commercialization. Our research suggests that despite the consensus building process industry organizers engaged in leading up to the formation of the NRI, the

overall level of public funding for beyond CMOS research in the years to come remained relatively constant. That said, by the time the NRI program was re-competed in 2013, its central tenet, the need to investigate computational devices built on alternative state variables, had become a guiding philosophy for public funding programs in the field broadly and industry leaders involved with the NRI had begun to shape the discussion about future trends in computational devices more broadly.

Prior to the launch of the NRI, the history of alternatives to silicon integrated circuits was marred by failed predictions and deliverables. Westinghouse's promise of a molecular electronics "breakthrough" in the 1960s never materialized (Choi & Mody 2009). Similarly, the promise of faster and more energy efficient computation from IBM's Josephson Junction computing project during the 1970s evaporated as technical and manufacturing challenges delayed the technology while silicon integrated circuits continued their unrelenting progress (Mody 2016). Moreover, molecular electronics research in the late 1990s and early 2000s associated with government nanoelectronics programs faced setbacks from a series of scientific scandals (Choi & Mody 2009). In addition to the history of failed alternatives, early evaluations of emerging technology research by ITRS working groupsin 2001 and 2003 concluded that none of the technologies examined was projected to surpass CMOS in every major metric (speed, energy efficiency, density, cost) with many failing to surpass CMOS in most. Furthermore, the research project which grew out of the ITRS ERD chapter's examination of emerging technologies concluded that many existing research streams were unlikely to deliver any significant improvements over ultimate CMOS due to fundamental limits of heat transfer for charge based devices (Zhirnov et al. 2003). Beginning in 2005, the ERD began issuing survey based assessments of emerging technologies, finding only one emerging logic technology (1D

Structures) to be potentially viable (ITRS 2005). While early evaluations of emerging technologies relied on expert projections of high-level attributes (e.g. "Scalability" or "CMOS compatibility"), NRI researchers soon presented more quantitative estimates of device comparisons.

As mentioned above, the direct influence of industry researchers on NRI-affiliated academics meant that researchers with backgrounds in fundamental science had to consider system design and manufacturing limitations for new device concepts. NRI industry leadership instituted a series of workshops and lectures for students and professors to educate them on "what needs to be done to make devices interesting to industry" (Marshall 2013). This process not only affected the types of technologies considered most promising but began to shape how academic researchers viewed the problem.

Evaluating the progress of NRI device concepts that used different computational state variables remained one of the NRI program's principal challenges. Early in the program, each center had researchers working separately on the issue of device metrics. Industry members are blunt that early self-reported device metrics were unreliable. This unreliability led to a formalization of the process with the introduction of NRI's benchmarking program in 2009, led by researchers from the MIND center and Kerry Bernstein, at the time of IBM. The benchmarking program attempted to compare devices being developed by different NRI centers using common metrics such as switching delay, energy, and device area for realized circuits based on new device types (Bernstein et al. 2010; Nikonov & Young 2013; Nikonov & Young 2015).[60] Researchers were asked to describe in detail how specific logic functions - e.g. an adder

[60] "To overthrow a conventional system frequently requires, in addition to the formulation of an alternative system and the definition of new performance parameters, the creation of new or much refined testing techniques" (Constant 1980 pg 22).

– would be realized by their proposed device and to provide quantitative estimates for key figures of merit. Figure 18 shows one of the initial benchmarking report's main outputs. Results from these reports were also published in the ITRS ERD chapters. Although those involved with the benchmarking caution that the accuracy of the results is limited because few devices have working prototypes, the process shaped both academic and industry research directions.[61] Tom Theis, the NRI's director through 2016, argued that the "benchmarking program has really helped people in the academic community to understand what the problem is" because it allows an apples-to-apples comparison. Similarly, Dimitri Nikonov, an Intel liaison and a leader of the benchmarking program, said the program allowed the industry "to set future research directions" by focusing efforts on what looked most promising, akin to the search processes that (Vincenti 1990) describes as selection and retention. While some academic participants worried initially that the benchmarking program would be used to kill projects, afterward they admitted that they had gained "a better understanding of the problems facing the industry," and they subsequently reshaped their device concepts to overcome initial limitations (Theis 2013).

In particular, whereas earlier estimates simply compared device characteristics to those of CMOS these results indicated potential new areas of focus for future research directions. The 2011 ITRS ERD summarized the takeaway as such, "several of the devices appear to offer advantage over CMOS in logical effort, particularly for more complex functions. These findings increase the urgency of doing more joint device – architecture co-design for these emerging technologies" (ITRS 2011). Similarly, a 2015 update by Intel researchers found that for a given emerging device technology interconnects were the largest contributors to switching energy and

[61] Here, too, Constant's observations are trenchant: "Use of prior technology, either directly or as a model or guide, does immensely simplify the search for problem or subproblem solution: it is highly efficient for design. Yet to the extent that a prior technology shapes the formulation of systems in a completely alien field, its use can only exacerbate an already abstruse decision process" (Constant 1980 pg 19).

delay (Nikonov & Young 2015). In both cases these results suggested new avenues (architecture and interconnects) for research to improve the feasibility of non-CMOS emerging device technologies beyond simply improving device-level characteristics.

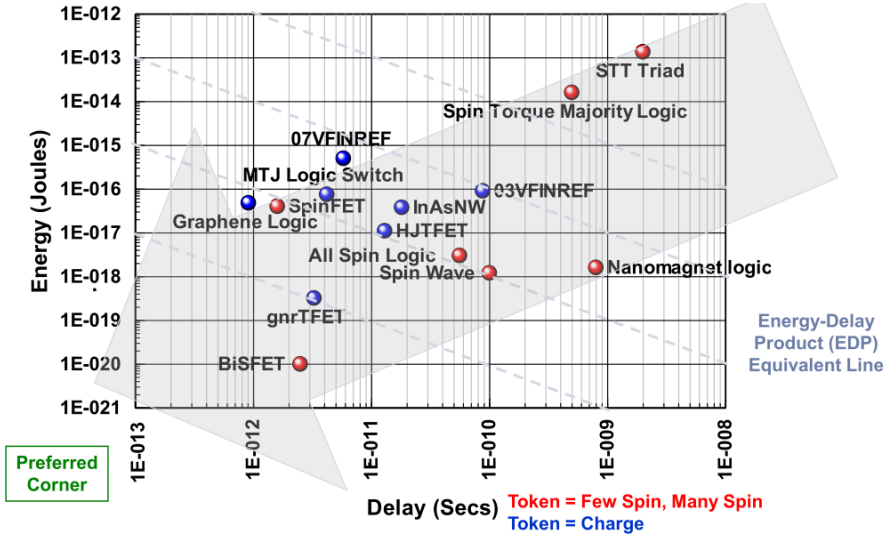

Figure 18 - NRI benchmarking results comparing switching speed and energy for NRI device concepts. Source: Bernstein et al., 2010

Similar to the role of the NRI in shepherding public funding investments, industry leadership from the NRI also influenced the wider research community. NRI assignees and liaisons constituted a special group of individuals who had interacted with each other through their firms' participation in other SRC and ITRS working groups. For example, several of the NRI liaisons and assignees also contributed to the ITRS ERD chapter and shared NRI research findings through that publication. The contributions to the ITRS, in turn, shaped the research directions of the wider community of researchers beyond the NRI. An examination of the technologies listed in the ITRS ERD chapter shows a significant overhaul of the technologies presented as most promising. Technologies considered by government programs in the 1990s were removed and new concepts, many being researched by the NRI, were added. These changes are summarized in table 1 in the previous chapter. In December 2013, NRI's industry representatives presented NRI benchmarking results at the industry's largest device conference,

IEDM. At the same conference, NRI industry representatives also presented a short-course on "Beyond-CMOS" devices and shared a paper summarizing the results of NRI's benchmarking study. NRI leadership noted that researchers working on neuromorphic devices and superconducting circuits – devices not in NRI's research portfolio –planned to adopt benchmarking metrics similar to NRI's process for their own devices (Nikonov 2013). Thus, NRI's industry representatives not only played a role in transferring NRI-developed technologies to their own firms but also in communicating new insights on technical directions and benchmarks across the larger scientific community.

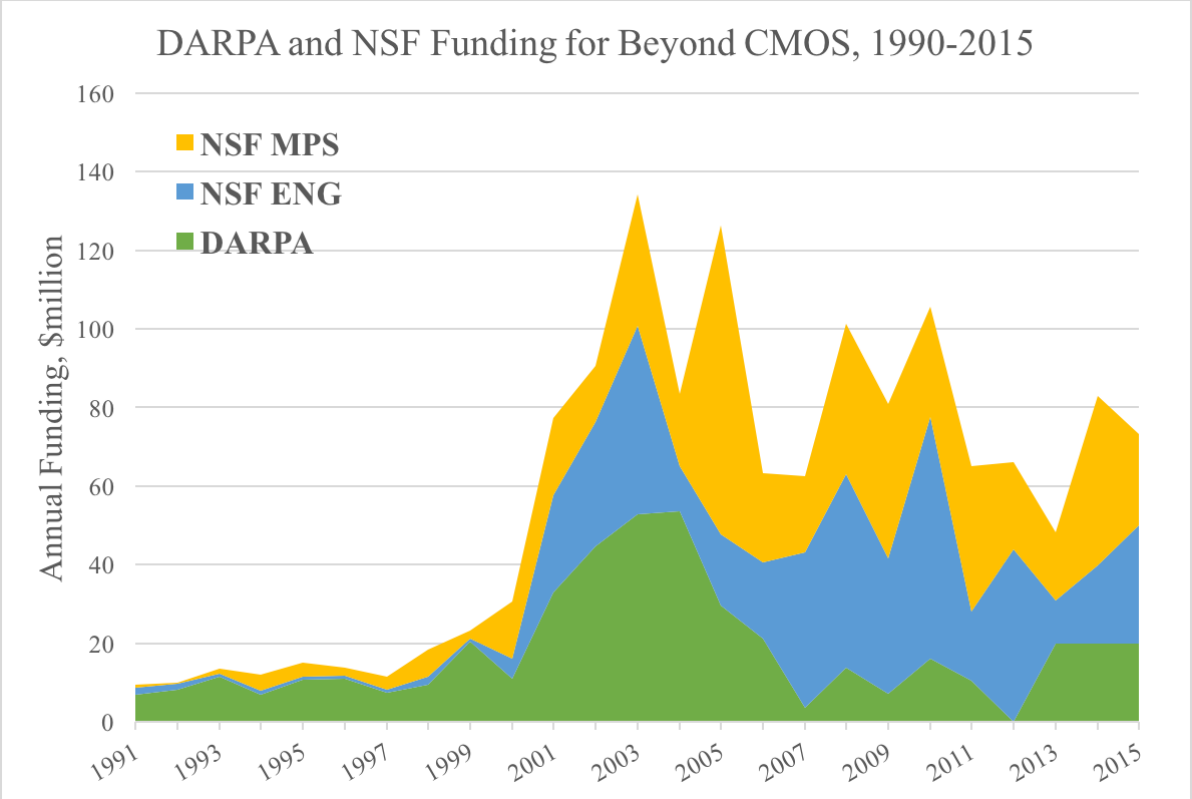

Figure 19 Total funding for beyond CMOS programs at DARPA and nanoelectronics programs at NSF in the Engineering and Materials and Physical Sciences Directorates 1990 – 2015. Source: (RDSS 2017; NSF 2017)

Despite these NRI-affiliated activities, there has been a distinct stagnation in overall public support for beyond CMOS devices in the United States since the inception of the NRI.

DARPA, in particular, ramped up funding for alternative computing technologies in the early-2000s prior to the launch of the NRI. Beginning in the early 2000s, DARPA began a series of programs looking at Spin electronics (2003-2006), Sub-threshold slope transistors (2008-2009), and Non-Volatile Logic (2010-2011) in addition to a host of programs researching quantum computing technologies, an area of increasing investment by defense related agencies. In 2005-2006, DARPA's collection of programs under its "Beyond Silicon" umbrella totaled $80 million in funding. However, by 2012, DARPA's focus had shifted toward quantum computing and the majority of its beyond CMOS funding in the years following was through the SRC's STARNet program. DARPA's molecular electronics programs – Moletronics and MoleApps – which grew out of its ULTRA program were phased out by 2007. Meanwhile, NSF funding for "beyond CMOS" and "nanoelectronics" research in its materials and engineering directorates has been fairly consistent in the decade and a half since the launch of the NNI. Importantly, the data is drawn from keyword searches of the NSF funding archives, so the massive uptick in the new millennium may be spurious as researchers began to recast their research efforts using buzzwords associated with the NNI (e.g. "nanoelectronics"). While not an exhaustive accounting of public funding efforts in the field, that overall nanoelectronics funding levels have stagnated – and likely declined – is corroborated by self-reported figures from agencies available in the latest NNI triennial review which reported an average of $90 million for nanoelectronics from 2011-2015 (National Academies of Sciences, Engineering 2016). This stagnation in funding comes despite an emerging consensus amongst researchers that beyond CMOS technologies may not be able to directly compete with CMOS and instead achieving long-term advantages over CMOS will require advances in emerging device technologies combined with innovations in interconnects and architecture (Nikonov & Young 2015).

There is also evidence of a growing influence of the industry on the structure and direction of publicly funded efforts in the field of nanoelectronics. Several of the leading organizers involved with the SNB workshops in the US also were instrumental in establishing internationally oriented groups: the International Nanotechnology Conference on Communication and Cooperation (INC) and the related International Planning Working Group on Nanoelectronics (IPWGN). These two groups continued to bring together high-level leadership from government, industry, and academia on defining nanoelectronics research directions. The IPWGN, for example, published reports surveying nanoelectronics funding in the US, Europe and Japan along the research vectors first defined at the SNB and adopted by the NRI (Brillouët et al. 2010; Roco 2011; Galatsis et al. 2015). In 2010, the National Nanotechnology Initiative issued a set of Nanotechnology Signature Initiatives (NSI). The "Nanoelectronics for 2020 and Beyond" (NEB) NSI listed five separate thrusts for its collaborating agencies, the first of which was "explorning new or alternative 'state variables,' architectures, and modes of operation for computing" (NSET 2010). The NEB's first research thrust was identical to the first research vector adopted as the core of the NRI's program. While industry may be leading the direction of research, representatives from each of the defense (DARPA, ONR, AFOSR, ARL) and civilian (NSF, DOE, NIST) research agencies have continuously attended NRI reviews. Thus, a dense web of interrelated groups – the NRI TPG, ITRS ERD, INC, IPWGN – drawing on core group of industry leaders, academics, and government officials have continued to shape national and international research directions in the field of nanoelectronics since the 1990s.

These relationships are summarized in table 5. ADT board includes individuals who were placed on the SRC's "Technical Advisory Board" for the ADT program in 1999/2000. ITRS

ERD lists contributors to the ERD chapter over its lifetime (2001-2013). "NNI-SRC WG" includes individuals placed on the interagency working groups following the first SNB workshops. INC includes individuals on the steering committee for the INC. NNI review includes individuals present at the 2004 "Nanoelectronics, Nanophotonics, and Nanomagnetics" Workshop that coincided with the second SNB workshop. SIA Nano TSC includes industry executives that served on the SIA's technology strategy committee. NRI TPG/GC includes individuals who served on the leadership boards of the NRI. NRI Centers indicates those with research-level contributions to the NRI.

Table 5 - Visual summary of key individuals that shaped SRC's "beyond CMOS" strategy from 1999 through launch of NRI.

Name	Organization	ADT Board	ITRS ERD	NNI-SRC WG	INC	NNI Review	SIA Nano TSC	NRI TPG/GC	NRI Centers
Tak Ning	IBM		X	X				X	
Tom Theis	IBM		X		X			X	
H.S. Philip Wong	IBM	X	X	X		X	X		
George Bourianoff	Intel/SRC	X	X	X	X	X	X	X	X
Paolo Gargini	Intel			X	X		X		
Dimitri Nikonov	Intel		X						X
Luan Tran	Micron		X				X		
Zoran Krivokapic	AMD/GF	X	X					X	
Ming-Ren Lin	AMD/GF	X						X	
Wilfried Haensch	IBM	X						X	
Bob Doering	TI			X	X		X	X	
Mihail Roco	NSF			X	X	X			
Gernot Pomrenke	DoD			X	X	X			
David Seiler	NIST			X	X			X	
Ralph Cavin	SRC			X	X				
Jim Hutchby	SRC	X	X	X		X	X		
Daniel Herr	SRC		X	X					
Victor Zhirnov	SRC		X						
Dimitri Antoniadis	MIT		X	X		X			X
Mark Lundstrom	Purdue		X	X		X			X
Kang Wang	UCLA		X	X	X	X			X
Year of Formation		1999	2001	2004	2004	2004	2004	2005	2006

4. Discussion: Institutional Failure at the End of a Technology Paradigm?

The case of the semiconductor industry at the end of Moore's Law may offer new insights for our understanding of the role and capabilities of institutions in shaping and

responding to challenges at the end of a technology paradigm. Scholars have argued that the National Nanotechnology Initiative is best understood as targeted industrial policy for the semiconductor industry (McCray 2005) and that firms are now less likely to support the type of basic research that has historically contributed to massive new business lines (Arora et al. 2015). Our research both critiques this understanding of the NNI and highlights industry engagement with scientific discovery that may be discounted by analysis of publication statistics.

Unlike those who argue that technology transitions are mediated by the intrinsic capabilities of a technology (Foster 1986) or the result of path-dependent processes that build on exogenous developments (David 1985; Arthur 1989), this study is in the tradition of scholars who have noted that social and organizational factors are particularly important during periods of technological uncertainty (Dosi 1982a; Constant 1987; Tushman & Rosenkopf 1992). We further contribute to this research by examining under what conditions existing institutions – markets and firms, public actors, universities and cooperative technical organizations – may fail to bridge to the next paradigm.

4.1. Institutions at the end of a Paradigm

4.1.1. Markets and market actors

This study makes several contributions to our understanding of the role of markets and market actors – i.e. firms – in addressing the end of a technology paradigm. While the literature on technology (Utterback 1994) and industry (Klepper 1996; Klepper 1997) life-cycles examines the evolution of market competition after the introduction of new technologies, we show how researchers and executives from within our focal firms actively worked to reshape the directions of scientific researchers well before any new technologies are commercialized. The engagement with the scientific community documented here also suggests a more nuanced use of external

science by firms than simply as an input to their own technologies and is distinct from the approach of developing a firm's own absorptive capacity (Cohen & Levinthal 1990). NRI's assignees and liaisons worked to shape research directions at several levels: involvement in NRI governing boards, direct research collaboration with academic researchers, and consultation and engagement with the wider industrial community and government funding agencies. These roles differ markedly from those played by industry employees in short-term-focused consortia (e.g. SEMATECH, MCC, VLSI), where employees primarily worked on managing and evaluating technical progress of development projects (Grindley et al. 1994). The NRI assignee and liason roles were also distinct from university-industry collaborations on individual projects where the focus was on bringing ideas back into the firm. In part because NRI's early research results were less directly appropriable by their employing firms, the assignees focused on improving research directions more broadly. Additionally, while a host of research suggests that existing firms struggle to cope with paradigm shifting technologies due to competence obselescence (Tushman & Anderson 1986; Henderson & Clark 1990; Abernathy & Clark 1985), our study indicates that only the types of vertically integrated manufacturers identifed by (Kapoor 2013) to have contributed systemic innovations in the industry over the past several decades are researching these technologies.

However, despite the leading role these firms have taken in shaping scientific research directions through the NRI, it is unclear that these firms will be able to successfully commercialize a beyond CMOS device. As a result of the decreasing returns from traditional methods of scaling transistors, distinct trajectories driven by rapidly changing product markets have emerged. The industry's largest customers have vertically integrated upstream into chip design and in many emerging product markets (e.g. machine learning, computer vision)

propietary accelerator chips provide orders of magnitude of improvement compared to that possible through traditional manufacturing advancements (i.e. transistor scaling). These developments suggest dynamics similar to those examined in previous studies of the photolithography and hard disk drive industries where existing technologies continued to persist past their perceived limits despite concentrated effort by firms to invent and commercialize new technologies (Christensen & Rosenbloom 1995; Henderson 1995; Adner & Kapoor 2016). Absent from NRI's research program were vertically specialized semiconductor firms, design and equipment supply firms, and downstream users of semiconductor products. This narrow membership limited the ability of the consortium to tackle the challenge of replacing CMOS at a more comprehensive system and architecture level.

In recent years, IBM has sold off its semiconductor manufacturing facilities and refocused its business on providing technology services. Meanwhile, Intel has commited tens of billions of dollars to new product markets through acquisitions. The ability of these firms to devote the considerable resources necessary to commercialize a beyond-CMOS device may depend in large part on the financial success of their new business strategies, which remains unknown. Furthermore, during the period from 2007 through 2011, venture capital investment in the industry dwindled with only 36 firms receiving funding, fewer than the 44 that received funding in 2003 alone, suggesting that private markets are not well suited to allocating the resources necessary to invent and commercialize a radically new device.

4.1.2. Governments and Public Funding Agencies

Previous research has argued that the institutions designed to guide science and technology are products of national competitiveness efforts (Nelson 1993) and, in the case of the

U.S., the result of political compromise (Hart 1998). As a result, concentrated public efforts in technology have primarily occurred for the purposes of defense or technology catch-up in industries deemed crucial for economic competitiveness. Defense agencies played numerous roles in supporting the early development of transistor and integrated circuit technologies (Levin 1982; Holbrook 1995; Lecuyer 2005). Similarly, the justification for SEMATECH was along both military and economic lines (Browning & Shetler 2000). However, the semiconductor industry at the end of Moore's Law does not neatly fall into either of these catogories, creating a situation in which government agencies lack a clear mandate or impetus for intervention despite increasing high-level awareness of looming technology and industry challenges (Platzer & Sargent Jr. 2016; PCAST 2017).

First, there is no clear-cut competitiveness deficit to address. Chipmakers all over the world face the same technical challenges and there is no indication that the public research programs of other countires in nanoelectronics are more advanced (Galatsis et al. 2015). Additionally, recent history of the semiconductor industry suggests that public funding may be hard to come by even if other nations ramp up their research programs. In the years following the end of federal support for SEMATECH continued public funding for follow-on collaborative efforts in the transition to 300 mm wafers (Ham et al. 1998) or the development of EUV lithography (Linden et al. 2000; Appleyard et al. 2008; Sydow et al. 2012) were hampered by political criticism of public funding flowing to internationally based suppliers. In the case of the transition to 300 mm wafers, Japanese firms formed a rival consortia with financial support from the Japanese government. Moreover, the dynamics highlighted in the markets section above also complicate the discussion of whether there will be enough political will to mount large scale

programs to overcome "Moore's Wall" if new innovations continue to reach the marketplace despite the slowing of Moore's Law.

Second, whereas previous examples of focused government intervention were led by the Department of Defense, defense agencies actually reduced their funding for beyond-CMOS research during the period after the launch of NRI. In that decade, DARPA shifted resources toward technologies such as quantum computing, which is anticipated to offer benefits in applications of military interest such as cryptography. This lack of a lead agency is also implicated in related efforts led primarily by public agencies. The most recent triennial review of the NNI suggests that related government-led efforts such as the Grand Challenge for Future Computing, National Strategic Computing Initaitive (NSCI 2016) and the Nanotechnology Signature Initiative in Nanoelectronics (NSET 2010) have suffered from the "absence of any dedicated funding or a lead agency" (National Academies of Sciences, Engineering 2016).

4.1.3. Collaborative Technical Organizations and Communities of Practice

This study sheds further light on the form and function collaborative organizations can take during periods of scientific and technological uncertainty. Previous research has shown how conferences in the case of the cochlear implants (Garud 2008; Garud & Rappa 1994) and cooperative technical organizations in the flight simulator industry (Rosenkopf & Tushman 1998) shaped technical directions during eras of ferment (Utterback & Abernathy 1975; Tushman & Anderson 1986). Unlike the previous studies, in which the collaborative organizations were responding to technological uncertainty that had emerged due to regulatory uncertainty, our research suggests a role for collaborative organizations in both creating and responding to scientific and technological uncertainty. The work of industry researchers with

existing collaborative organizations – the SRC and ITRS – created scientific and technological uncertainty after researchers concluded that existing research efforts in emerging devices were unlikely to overcome the presumptive anomaly (Constant 1980) they had identified, i.e. limits to heat dissipation from charge based devices (Zhirnov et al. 2003).

Previous scholars have identified the importance of testing and technology evaluation as key processes in the emergence and acceptance of new technologies (Constant 1980; Mackenzie 1990; Vincenti 1990). NRI leadership, through its benchmarking and training programs codified variation, testing and retention processes (Vincenti 1990). This program combined with the NRI's influence on broader science funding programs helped create a practitioner community that shared in its procedures, traditions, and problem-solving modes (Constant 1987).

Another key difference in our research setting are the dynamics of collaborative organization creation and membership. In the case of DUV lithography (Kapoor & McGrath 2014), firms engaged with existing research consortia on science based collaborations during a technology's emerging phase. In the case of flight simulators, Rosenkopf & Tushman (1998) showed increased entry into CTOs during eras of ferment and suggest that "key" CTOs were founded during this period as well. In our setting, industry researchers and executives created an entirely new organization to address their specific scientific uncertainty and concomitant technology needs. However, whereas Rosenkopf & Tushman (1998) suggested individuals joined CTOs to strategically shape technology directions, in our research setting existing members of the community purposefully expanded membership to include new groups of researchers. This suggests that they dynamics of entry by individuals into collaborative organizations are mediated by the type of uncertainty – i.e. scientific versus technological - and whether the existing community has the appropriate expertise.

Furthermore, the question of whether the organizations founded during this period will emerge as "key" organizations as in the case of flight simulators remains open. The organizations created during this period, such as the NRI and IPWGN, leveraged much of their institutional support from already existing groups such as the SRC, SEMATECH, and the ITRS. In recent years as the industry underwent consolidation and new trajectories emerged these cooperative technical organizations have begun to falter. SEMATECH, the manufacturing consortium established in 1987, ceased standalone operations in 2015 after its largest members pulled out. The ITRS was also disbanded in 2015 after organizers noted that the industry was no longer driven by a single trajectory. A similar effort, the International Roadmap of Devices and Systems (IRDS), is now taking place under the auspices of the IEEE. Unlike the ITRS, which projected the evolution of commodity products, the IRDS will focus on the evolution of technical needs for different product markets and different emerging computing architectures and systems. Further, the number of firms paying membership dues to the SRC has been declining since the 1990s as many of the new vertically disintegrated entrants forewent membership in the consortium. These dynamics are further exacerbated by the fact that many of the industry's customers that are now vertically integrating upstream into chip design are not members of the industry's longstanding collaborative organizations, and are not expressing interest in becoming members thereof. In contrast, at the founding of the SRC, the industry's then largest customers, computer manufacturers, were amongst the consortium's earliest members.

Conclusions and Policy Issues at the End of a Paradigm

This chapter discusses the various issues that policymakers must consider in evaluating different options as the semiconductor industry grapples with the maturation of the silicon integrated circuit paradigm. First, I discuss briefly the economic importance of semiconductors and why the slowdown in advancement of semiconductors is a significant issue economic and policy issue. Second, I briefly review the technical and institutional challenges facing the industry. Included in this section is a discussion of what is currently being attempted by industry as firms attempt to take advantage of new market opportunities. Then, I review existing policy efforts of the last several years and add new recommendations for policymakers.

1. The Importance of Semiconductors

The integrated circuit (IC) provides a superb example of what economists call a "general purpose technology"—a technology that makes possible other important technologies, products, and services, which in turn contribute importantly to economic growth and welfare. Owing to its long-term, systematic declines in cost and lockstep increases in performance (a.k.a., "Moore's Law"), the IC made possible a dazzling array of new or advanced products, from intercontinental ballistic missiles to global environmental monitoring systems and from smart phones to medical implants. Indeed, so fecund and important has the IC been that economists identified it as the "foundation for the American growth resurgence" in the 1990s and the leading source of worldwide economic growth during that same period. (Jorgenson 2001; Jorgenson & Vu 2007).

However, as documented in previous chapters, the semiconductor industry has not been able to maintain its historical rates of improvement. Beginning in the mid-2000s, gains from

geometric scaling were limited by power-density limits that resulted from an increase in passive power consumption (Haensch, E. J. Nowak, et al. 2006). It is evident from public announcements by semiconductor firms, as well as published ITRS roadmap projections (see figure 6 chapter 2), that the steep performance wall for microprocessors was not widely anticipated (Thompson 2017; National Research Council 2011). As a result, semiconductor manufacturers hastily adopted a multi-core approach. Recent evidence suggests that end-users were unable to take advantage of the performance potential of multi-core processors the way they had previous advancements in semiconductors due to the difficulty of parallel programming approaches (Thompson 2017; National Research Council 2011) and that this inability to parallelize applications may explain an overall decline in the contribution of IT-using sectors to total factor productivity growth (Thompson 2017; Ho et al. 2011). The contribution of semiconductors to total factor productivity growth during the 1990s and the subsequent slowdown after the introduction of multi-core processors both highlight the importance of semiconductor technology to overall economic growth.

2. Technical and Institutional Challenges at the End of Moore's Law

Public funding played a crucial role in the establishment of both computing and semiconductor technologies (Flamm 1987; Flamm 1988; Riordan & Hoddeson 1997; Holbrook 1995; National Research Council 1999). However, by the early1980s the semiconductor industry explicitly noted that federal funding for semiconductor research at universities, which came primarily through the Department of Defense, was increasingly unrelated to commercial technology challenges (SRC 1983). In addition to the disparate needs of federal defense agencies and semiconductor manufacturers, new international competition from Japan and the first phase

of the industry's vertical disintegration put pressure on the industry's long-standing research and development model. Some of industry's largest centralized research laboratories either ceased to exist or scaled back their investments in long-term research, and the industry responded by creating a set of institutions that embraced collaborative research and direction setting. In the case of the Semiconductor Research Corporation (SRC), the industry acted unilaterally from government agencies in setting up an alternative funding pool for academic research derived solely from industry funds focused on silicon integrated circuits and specifically CMOS (Sumney & Burger 1986). Meanwhile, the creation of SEMATECH and the National Advisory Committee on Semiconductors were done in conjunction with the federal government and motivated by the industry's importance to national security and economic competitiveness. Through the 1990s and 2000s, this group of organizations – the SRC, SEMATECH, and NTRS (later, ITRS) – coordinated the research efforts of government agencies, academic researchers, industry suppliers, and manufacturers (Schaller 2004; Gargini 2000). By the late 1990s, the national competiveness charter of the industry's collaborative institutions gave way to a focus on maintaining Moore's Law as new technical hurdles emerged. As a result, these organizations began to admit international members.

2.1. New Challenges at the End of a Trajectory

Today, however, the industry faces its most daunting technical and institutional challenges. Although semiconductor manufacturers have introduced materials and process innovations to continue "equivalent scaling," recent evidence suggests that the historical rate of cost decreases and performance improvements that came with transistor scaling have ceased (Hruska 2012; Eassa 2017). Intel announced a departure from its historic "tick-tock" strategy in 2016, effectively conceding that the cadence of Moore's Law had slowed to a three year pattern

from two years (Cutress 2016). A significant contribution to the slowdown in cost-per-transistor decreases has been the decade-long delay of the scale-up and introduction of extreme ultraviolet (EUV) lithography into production processes. While leading manufacturers have begun to include the technology on product roadmaps, technical hurdles remain (Lepadus 2017). In its stead, firms have relied on double and quadruple patterning with existing 193nm lithography, which has driven up process complexity and associated costs. Regardless, current projections indicate that the successful introduction of EUV lithography will likely only provide a temporary reprieve from escalating costs (Sperling 2017).

Additionally, the institutions that guided the semiconductor research ecosystem have begun to falter. The SRC has seen its number of member firms decline consistently since the 1990s and has struggled to recruit specialized design (fabless) firms. SEMATECH ceased to exist as a standalone entity in 2015 and was folded into the SUNY Albany College of Nanoscience. Similarly, the SRC and SEMATECH stopped contributing to the ITRS in 2013, and the SIA pulled support in 2015. The decline of the industry's collaborative research institutions is a consequence of industry-wide consolidation and a fundamental shift in the industry's rate and nature of technological progress. There has been a rapid drop in the number of manufacturers operating at the newest process nodes since 2000. Thus, much of the research funded by organizations such as SRC, SEMATECH, and IMEC (the European consortium located in Belgium) is only of use to a dwindling number of firms. Although both the SRC and IMEC have responded by creating "customization" programs where firms can target specific research projects with their contributions, the collaborative nature of these organizations has been hampered. In the case of EUV, with only a handful of firms operating at the newest process nodes, vertical collaboration through SEMATECH with equipment manufacturers has been

replaced by direct equity investments; Samsung, TSMC, and Intel all invested directly in the leading EUV manufacturer, ASML (Intel Corporation 2012).

As noted above, the fraying of the industry's collaborative research structure is linked to the industry's departure of its historical model of technological progress where advances in CMOS technology drove product improvements in CPUs and DRAMs that underpinned advances in computing technology. Other firms in the industry also benefitted from the advances in manufacturing technology made possible by these product markets. Under this technological regime, the semiconductor industry's largest customers were supportive of the formation of the industry's collaborative institutions. For example, computer manufacturers such as IBM, DEC, Honeywell, and Hewlett-Packard were early members of the SRC. However, in recent years, as the nature of technological progress in the industry has evolved from a single trajectory to multiple trajectories differentiated by product and market needs, the incentives of customers and manufacturers have also splintered. One example is the decision by large technology firms to vertically integrate upstream into chip design. As the gains from traditional scaling methods have slowed, firms have been able to deliver orders of magnitude improvement on relevant performance metrics with the use of application specific integrated circuits (Horowitz 2014). For example, in announcing its Tensor Processing Unit (TPU) designed for machine learning applications, Google claimed 100x improvements in specific performance metrics over commercially available commodity chips despite using older manufacturing technology (Jouppi et al. 2017). Furthermore, researchers also argue that there continues to be much room for improvement "at the top" in areas such as software design and computing architecture to improve overall system performance (Theis and Wong, 2017). These arguments are further corroborated by the pace of advancement in fields such as machine learning, computer vision,

and robotics, which has far outstripped that of the underlying semiconductor technology each of them uses. As these and other emerging technologies continue to improve and develop their own mature ecosystems, their advancement may become further decoupled from that of semiconductor manufacturing technology.

2.2. Splintering Trajectories in a "Post-Moore" World

As a result, in place of the ITRS, a new effort, the International Roadmap for Devices and Systems (IRDS), spearheaded by the IEEE's Rebooting Computing Initiative, began in 2016. While the ITRS was built around the technical drivers for CMOS in the semiconductor industry's two largest standard products, DRAM and CPUs, the IRDS is built around the notion that there will be distinct technical drivers for different product markets possibly on different technology platforms. Table 6 and Figure 20 below summarize the current IRDS schema for emerging products, devices, and architectures (IRDS 2016). While these are clearly in their early stages, it is evident that there is significant uncertainty over the technological capabilities of emerging technologies and their ability to address disparate market needs.

The IRDS projections are also a stark departure from the industry's earlier stated goals of identifying a successor to CMOS. The language used by industry leadership around the founding of the NRI was explicitly geared toward maintaining the Moore's Law trajectory. For example, Welser et al. (2008) argued "[a] new 'switch' for information processing is needed to significantly extend the scaling path." Other publications by NRI associated researchers argued for a "paradigm shift" (Bourianoff and Theis, 2004) or the need to "reinvent the transistor" (Theis & Solomon 2010). The language being used by the IRDS now recognizes that a paradigm shift that maintains the existing trajectory may not be possible and that instead product markets

will utilize new technologies in a variety of ways, i.e. there is a splintering of trajectories as the existing paradigm ends.

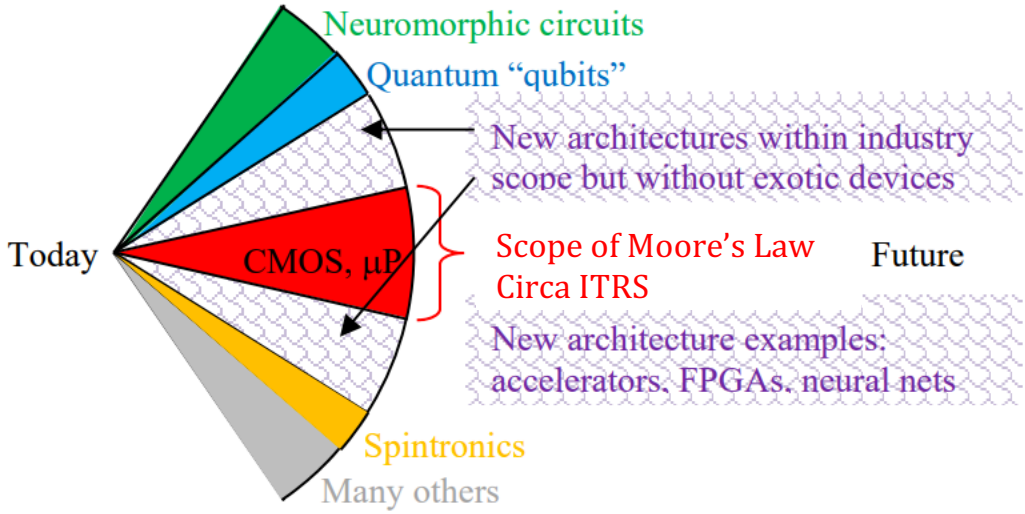

Figure 20 - Current IRDS schema for emerging semiconductor trajectories.

The compounding scientific, technological, and market uncertainties over the future of new technologies raise questions about the ability of the semiconductor industry's remaining institutions to successfully introduce an alternative to CMOS. A commercially successful alternative to CMOS will require reducing to practice an entirely new computing element based on a new materials system that likely operates using a different set of physical phenomena than traditionally relied upon by the industry. Commercial success of such a technology will require, eventually, new manufacturing techniques and design tools to bring these devices to market. Yet, the institutions that historically shaped the evolution of basic research into radically new technologies – corporate research programs and military demand - are greatly weakened. In their stead, the industry has developed institutions that coordinated networks of academic and industry researchers along the industry's pre-defined trajectory (Mody 2016). However, given the new nature of technological change underway in the semiconductor and computing industries, they

may be hampered by their relatively limited membership. As highlighted in the move from single- to multi-core processors, downstream technology shifts require coordination up and down the computing technology stack to be effectively employed. While narrowly focused public-private partnerships have been successful in the industry's past, the current paltry response to the end of Moore's Law raises fundamental questions about the existing institutions' appropriateness for an effort with this degree of scientific uncertainty and technological complexity.

Table 6 - Current IRDS schema for product market and technology drivers. Source: IRDS, 2016

Market drivers	**Technology focus**	**Scaling focus**	**Specific technologies**
Mobile Driver	Form factor Power, thermal Performance Features	Cost	Multi-Vt 3D integration Unified logic/RF Advanced DRAM (HBM, HMC) Integrated multi-standard comm. Circuit Sensor/MEMS/logic integration
Microserver Driver	Latency Bandwidth Off-MPU bandwidth	IO device Leakage 2.5D/3D integration Memory BW/latency	DRAM, eDRAM, MRAM, RRAM Silicon photonics (incl. silicon compatible laser source) High-radix networks Distributed compression/encryption engines Novel memory devices 64-bit ARM core Modularized 3D stacks
Datacenter Driver	Latency Bandwidth Off-MPU bandwidth Power density	Performance Reliability Memory BW/latency	
IoT Driver	Form factor Power, thermal Energy harvesting	Specialty devices at baseline tech SiP miniaturization	Vt control eFLASH, HV, MRAM, RRAM 3D integration On-chip passive components Configurable/fine-grain regulation New computation paradigm (near-threshold, asynchronous, stochastic, approximate)

3. A Review of Current Policy Efforts and Additional Recommendations

Many of the ongoing research and funding programs in the field of beyond CMOS devices were covered in chapters 2 and 3. DARPA began a series of programs in beyond-CMOS technologies around the turn of the century, building on the ULTRA program. By 2007-2008, most of those programs had been phased out, and much of DARPA's funding in beyond CMOS space in the years since has been through the SRC's FCRP and STARnet programs. The most notable research program in the field of beyond CMOS has been the Nanoelectronics Research Initiative (NRI), first founded in 2005. Federal funding of NRI has come through both NSF and NIST, with overall funding levels averaging roughly $20 million annually since inception. Beginning in 2010, the National Nanotechnology Initiative organized a series of "Nanotechnology Strategic Initiatives" (NSI) of which "Nanoelectronics for 2020 and Beyond" was one. The NNI Triennial Review by the National Academies is the only source of published data on funding levels for the NSI since their announcement, and it shows an average funding of $86 million for nanoelectronics NSI since 2010, with projected declines in 2016 and 2017 (National Academies of Sciences, Engineering 2016).

In 2017, the President's Council of Advisors for Science and Technology (under Barack Obama) issued a report titled, "Ensuring Long-Term US Leadership in Semiconductors." It centers on the threat of commodification posed by the significant state-backed investments in semiconductor manufacturing technology being made in China and suggests a series of "moonshots" sponsored by the US government that would simultaneously address federal agency missions and stimulate the development of advanced semiconductor technology (PCAST 2017). However, one such moonshot was announced in 2015 by the Obama administration: the "Nanotechnology Inspired Grand Challenge for Future Computing" remains the sole grand

challenge announced by the NNI. Meanwhile, the latest Triennial Review of the NNI issued by the National Academies has criticized the grand challenge and NSI approaches for their lack of lead agencies, dedicated funding, or appropriate expertise on the relevant committees that determine the statements of need (National Academies of Sciences, Engineering 2016).

The array of scientific, technological, and commercial possibilities for future semiconductor and computing technologies presents a significant challenge to scientists, engineers, and policymakers alike. The magnitude of the social benefits to be gained through a post-CMOS general-purpose technology that will scale in the manner of Moore's Law, coupled with the substantial basic research and platform coordination required to move beyond the current technology paradigm, suggests that a much more substantial public funding and coordinating initiative is required than is currently in place or anticipated. Furthermore, the scale of the scientific and technological challenge puts the onus on policymakers to think carefully about the proper organizational form for allocating public funds. Policymakers should resist the urge to consolidate the public's research portfolio and instead focus on cultivating numerous parallel approaches to beyond CMOS (Nelson 1961; Scherer 2011). Despite high-level awareness of the industry's technology challenges and opportunities (National Academies of Sciences, Engineering 2016; Holdren & Donovan 2016; PCAST 2017), the recommendations coming out of these earlier reports and studies have seen little urgent follow-through.

Given the uncertainties and rapid pace of advancement in various emerging technologies, it is unlikely that government "moonshot" initiatives will generate sufficient interest to drive industry effort. Instead, government agencies should focus on two broad objectives. First, increasing overall support for engineering and physical science research funding is imperative. The data collected from the NSF and DARPA archives as well as the NNI Triennial Review

findings on NSI funding indicate an overall stagnation in support for nanoelectronics-related basic research. This stagnation is not an outlier, however. According to NSF data, average annual growth in funding for engineering has averaged less than 1% and less than 0.5% for physical sciences since the launch of the NNI (National Center for Science and Engineering Statistics 2017). Second, federally funded consortia or public-private partnerships aimed at developing solutions for beyond CMOS should include firms from across the technology stack, not just semiconductor firms. While semiconductor manufacturers drove Moore's Law, the future of computing beyond the end of Moore's Law will require innovation at all levels of the technology hierarchy.

An important consideration for policymakers is where the locus of expertise to address the current technological discontinuity resides. Previous research has suggested that although the semiconductor industry has globalized its production, its research and development spending remains non-globalized (Macher et al. 2007) and that firms have largely stopped investing in basic research in-house (Arora et al. 2015). Meanwhile the findings of chapter 3 suggest that industry experts continue to play a pivotal role in shaping the research direction of academic scientists. Given these stylized facts, a robust science and engineering base in US universities may be the crucial input to generating viable concepts for beyond CMOS devices.

It is unclear whether the new administration intends to make post-Moore's Law R&D a priority. Advances in fields such as big data, robotics, and artificial intelligence hold considerable promise and social benefits. Yet each of them – to varying extents – relies on continued advances in scalable computing hardware. Successful commercialization of a "Beyond CMOS" technology may thus underpin global economic and military leadership in an array of emerging technologies. National and global strategies should unequivocally scale with the

avoided economic and social costs inherent in not successfully overcoming the end of Moore's Law and the upside gains in discovering and developing a robust replacement for the CMOS paradigm. Scaling Moore's wall is truly is one of the "grand challenges" of our time.

References

Abernathy, W.J. & Clark, K.B., 1985. Innovation: Mapping the winds of creative destruction. *Research Policy*, 14(1), pp.3–22.

Adner, R. & Kapoor, R., 2016. Innovation ecosystems and the pace of substitution: Re-examining technology S-curves. *Strategic Management Journal*, 37(4), pp.625–648.

Anacker, W., 1980. Josephson Computer Technology: An IBM Research Project. *IBM Journal of Research and Development*, 24(2), pp.107–112.

Ancona, D.G. (Deborah G. & Bresman, H., 2007. *X-teams : how to build teams that lead, innovate, and succeed*, Harvard Business School Press.

Anderson, P. & Tushman, M.L., 1990. Technological Discontinuities and Dominant Designs: A Cyclical Model of Technological Change. *Administrative Science Quarterly*, 35(4), p.604.

Angel, D.P., 1990. New Firm Formation in the Semiconductor Industry: Elements of a Flexible Manufacturing System. *Regional Studies*, 24(3), pp.211–221.

Anon, 2005. Darpa ends litho aid at critical juncture for maskless. *EE Times*. Available at: http://www.eetimes.com/document.asp?doc_id=1152833 [Accessed February 7, 2016].

Anon, 2004. Intel cancels Tejas, moves to dual-core designs. *EE Times*. Available at: http://www.eetimes.com/document.asp?doc_id=1150169 [Accessed December 8, 2015].

Appleyard, M.M. et al., 2008. The innovator's non-dilemma: the case of next-generation lithography. *Managerial and Decision Economics*, 29(5), pp.407–423.

Arora, A., Belenzon, S. & Patacconi, A., 2015. *Killing the Golden Goose? The Decline of Science in Corporate R&D [working paper]*,

Arrow, K., 1962. Economic welfare and the allocation of resources for invention. In NBER, ed. *The Rate and Direction of Inventive Activity: Economic and Social Factors*. Princeton University Press, pp. 609–626.

Arthur, W.B., 1989. Competing Technologies, Increasing Returns, and Lock-In by Historical Events. *The Economic Journal*, 99(394), p.116.

Assignee, 2013. Interview with Industry Assignee #1 in person at NRI Annual Review on October 22, 2013.

Baccarani, G., Wordeman, M.R. & Dennard, R.H., 1984. Generalized scaling theory and its application to a ¼ micrometer MOSFET design. *Electron Devices, IEEE Transactions on*, 31(4), pp.452–462.

Baldwin, C.Y. & Clark, K.B., 2000. *Design Rules: The power of modularity*, MIT Press.

Bassett, R.K., 2002. *To the digital age : research labs, start-up companies, and the rise of MOS technology*, Johns Hopkins University Press,.

Bate, R.T., 1990. Nanoelectronics. *Nanotechnology*, 1(1), p.1.

Bate, R.T. et al., 1987. Prospects For Quantum Integrated Circuits. In *Proceedings of SPIE Symposium on Quantum Well and Superlattice Physics*. pp. 26–35.

Bate, R.T., 1986. The future of microstructure technology — The industry view. *Superlattices and Microstructures*, 2(1), pp.9–11.

Bates, R.H. et al., 1998. *Analytic narratives*, Princeton University Press.

Bernstein, K. et al., 2010. Device and Architecture Outlook for Beyond CMOS Switches. *Proceedings of the IEEE*, 98(12), pp.2169–2184.

Bodway, G. et al., 1995. *SRC White Paper on Research Investment Gap Analysis*, Research Triangle Park, NC.

Bourianoff, G., 2006. *NRI History, Scope and Vision*,

Bourianoff, G., 2004. NRI research program planning R2. [Presentation to SIA Technology Strategy Committee August 2004].

Bourianoff, G. et al., 2006. SNB and CWG outputs and continuing activities [Presentation to SNB Workshops. September, 2006].

Bourianoff, G. & Theis, T., 2004. *NRI Motivation, Vision and Proposed Plan - Rev. 1*,

Bresnahan, T.F. & Trajtenberg, M., 1995. General purpose technologies "Engines of growth"? *Journal of Econometrics*.

Brillouët, M. et al., 2010. Regional, National, and International Nanoelectronics Research Programs: Topical Concentration and Gaps. *Proceedings of the IEEE*, 98(12).

Brock, D.C. & Laws, D.A., 2012. The early history of microcircuitry: An overview. *IEEE Annals of the History of Computing*, 34(1), pp.7–19.

Brock, D.C. & Moore, G.E., 2006. *Understanding Moore's law: four decades of innovation*, Chemical Heritage Foundation.

Browning, L.D. & Shetler, J.C., 2000. *Sematech: saving the U.S. semiconductor industry*, Texas A&M University Press.

Burger, R.M., 2000. *Cooperative Research: The New Paradigm [Available at: https://www.src.org/about/p001960.pdf]*,

Capasso, F. et al., 1989. Quantum functional devices: resonant-tunneling transistors, circuits with reduced complexity, and multiple valued logic. *IEEE Transactions on Electron Devices*, 36(10), pp.2065–2082.

Carayannis, E.G. & Alexander, J., 2004. Strategy, structure, and performance issues of precompetitive R&D consortia: insights and lessons learned from SEMATECH. *Engineering Management, IEEE Transactions on*, 51(2), pp.226–232.

Cavin, R.K. & Zhirnov, V.V., 2004. Silicon Nanoelectronics and Beyond: Reflections from a Semiconductor Industry–government workshop. *Journal of Nanoparticle Research*, 6(2/3), pp.137–147.

Cavin III, R., Herr, D.J.C. & Zhirnov, V. V., 2000. Semiconductor Research Needs in the Nanoscale Physical Sciences: A Semiconductor Research Corporation Working Paper. *Journal of Nanoparticle Research*, 2(3), pp.213–235.

Ceruzzi, P.E., 2003. *A history of modern computing*, MIT press.

Chatterjee, P.K., Ping Yang, B.S. & Hisashi Shichijo, B.E., 1983. Modelling of small MOS devices and device limits. *IEE Proceedings I Solid State and Electron Devices*, 130(3), p.105.

Cherington, P.W., Peck, M. & Scherer, F., 1962. Organization and Research and Development Decision Making Within a Government Department. In *The Rate and Direction of Inventive Activity: Economic and Social Factors*. Princeton University Press, pp. 395–408.

Chesbrough, H., 2003. Towards a dynamics of modularity: a cyclical model of technical advance. *The business of systems integration*, 174, p.181.

Chesbrough, H. & Kusunoki, K., 2001. The modularity trap: innovation, technology phase shifts, and the resulting limits of virtual organizations. In I. Nonaka & D. J. Teece, eds. *Managing Industrial Knowledge: Creation, Transfer and Utilization*. Sage Press London, UK, pp. 202–230.

Chesbrough, H.W., 2003. *Open innovation: the new imperative for creating and profiting from technology*, Boston, Mass.: Harvard Business School Press.

Chih-Tang, S., 1988. Evolution of the MOS transistor-from conception to VLSI. *Proceedings of the IEEE*, 76(10), pp.1280–1326.

Choi, H. & Mody, C.C.M., 2009. The Long History of Molecular Electronics Microelectronics Origins of Nanotechnology. *Social Studies of Science*, 39(1), pp.11–50.

Christensen, C.M., 1992a. EXPLORING THE LIMITS OF THE TECHNOLOGY S-CURVE. PART I: COMPONENT TECHNOLOGIES. *Production and Operations Management*, 1(4), pp.334–357.

Christensen, C.M., 1992b. EXPLORING THE LIMITS OF THE TECHNOLOGY S-CURVE. PART II: ARCHITECTURAL TECHNOLOGIES. *Production and Operations Management*, 1(4), pp.358–366.

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

Christensen, C.M. & Rosenbloom, R.S., 1995. Explaining the attacker's advantage: Technological paradigms, organizational dynamics, and the value network. *Research Policy*, 24(2), pp.233–257.

Cohen, W.M., 2010. Fifty years of empirical studies of innovative activity and performance Bronwyn H. Hall and Nathan Rosenberg, ed. *Handbook of the Economics of Innovation*, 1(1 C), pp.129–213.

Cohen, W.M. & Klepper, S., 1996. Firm Size and the Nature of Innovation within Industries: The Case of Process and Product R&D. *The Review of Economics and Statistics*, 78(2), pp.232–243.

Cohen, W.M. & Levinthal, D.A., 1990. Absorptive Capacity: A New Perspective on Learning and Innovation. *Administrative Science Quarterly*, 35(1), pp.128–152.

Cohen, W.M., Nelson, R.R. & Walsh, J.P., 2002. Links and Impacts: The Influence of Public Research on Industrial R&D. *Management Science*, 48(1), pp.1–23.

Constant, E.W., 1980. *The origins of the turbojet revolution*, Johns Hopkins University Press.

Constant, E.W., 1987. The Social Locus of Technological Practice: Community, System, or Organization. In *The social construction of technological systems: New directions in the sociology and history of technology*. Cambridge, MA: MIT Press, pp. 233–242.

Cooper, L., 2007. Another View of Nanoelectronics and Its Applications. [Presentation to NSF Nanoelectronics Group on October, 16, 2007].

Crane, D., 1972. *Invisible colleges; diffusion of knowledge in scientific communities.*, University of Chicago Press.

Culpan, T., 2016. Be a Rebel, TSMC. *Bloomberg*.

Cutress, I., 2016. Intel's "Tick-Tock" Seemingly Dead, Becomes "Process-Architecture-Optimization." *AnandTech*.

David, P.A., 1985. Clio and the Economics of QWERTY. *The American Economic Review*, 75(2), pp.332–337.

Dennard, R.H. et al., 1974. Design of ion-implanted MOSFET's with very small physical dimensions. *Solid-State Circuits, IEEE Journal of*, 9(5), pp.256–268.

Dennard, R.H., 1983. Physical limits to VLSI technology using silicon MOSFET's. *Physica B+ C*, 117, pp.39–43.

Dennard, R.H. & Wordeman, M.R., 1985. MOSFET miniaturization—From one micron to the limits. *Physica B+ C*, 129(1), pp.3–15.

Dosi, G., 1982a. Technological paradigms and technological trajectories. *Research Policy*, 11(3), pp.147–162.

Dosi, G., 1982b. Technological paradigms and technological trajectories: A suggested interpretation of the determinants and directions of technical change. *Research Policy*, 11(3), pp.147–162.

Early, J.M., 1959. Maximum rapidly-switchable power density in junction triodes. *IRE Transactions on Electron Devices*, 6(3), pp.322–325.

Eassa, A., 2017. The Price of Intel Corporation's 10-Nanometer Failure. *The Motley Fool*.

Edgar, L.J., 1930. Method and apparatus for controlling electric currents.

Eisenhardt, K.M., 1989. Building Theories from Case Study Research. *Academy of Management Review*, 14(4), pp.532–550.

Flamm, K., 1988. *Creating the Computer: Government, Industry, and High Technology*, Brookings Inst Pr.

Flamm, K., 1987. *Targeting the Computer: Government Support and International Competition*, Brookings Inst Pr.

Fontana, R. & Malerba, F., 2010. Demand as a source of entry and the survival of new semiconductor firms. *Industrial and Corporate Change*, 19(5), pp.1629–1654.

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

Freeman, C., 1991. Networks of innovators: A synthesis of research issues. *Research Policy*, 20(5), pp.499–514.

Fuchs, E.R.H., 2010. Rethinking the role of the state in technology development: DARPA and the case for embedded network governance. *Research Policy*, 39(9), pp.1133–1147.

Galatsis, K. et al., 2015. Nanoelectronics Research Gaps and Recommendations: A Report from the International Planning Working Group on Nanoelectronics (IPWGN) [Commentary]. *IEEE Technology and Society Magazine*, 34(2), pp.21–30.

Gargini, P., 2004. Industrial Research Institute-Concept [Presentation to SIA TSC, January 14 2004].

Gargini, P., 2000. The International Technology Roadmap for Semiconductors (ITRS): "Past, present and future." In *GaAs IC Symposium. IEEE Gallium Arsenide Integrated Circuits Symposium. 22nd Annual Technical Digest 2000. (Cat. No.00CH37084)*. IEEE, pp. 3–5.

Garud, R., 2008. Conferences as Venues for the Configuration of Emerging Organizational Fields: The Case of Cochlear Implants. *Journal of Management Studies*, 45(6), pp.1061–1088.

Garud, R. & Rappa, M.A., 1994. A Socio-Cognitive Model of Technology Evolution: The Case of Cochlear Implants. *Organization Science*, 5(3), pp.344–362.

Gawer, A. & Cusumano, M.A., 2002. *Platform leadership*,

Gilbert, R., 2006. Looking for Mr. Schumpeter: Where Are We in the Competition--Innovation Debate? *Innovation Policy and the Economy*, 6, pp.159–215.

Glaser, B.G. & Strauss, A.L., 2012. *The discovery of grounded theory: strategies for qualitative research*, New Brunswick, N.J.: Aldine Transaction.

Goldey, J. & Ryder, R., 1963. Are transistors approaching their maximum capabilities? In *1963 IEEE International Solid-State Circuits Conference. Digest of Technical Papers*. Institute of Electrical and Electronics Engineers, pp. 20–21.

Grindley, P., Mowery, D.C. & Silverman, B., 1994. SEMATECH and collaborative research: Lessons in the design of high-technology consortia. *Journal of Policy Analysis and Management*, 13(4), pp.723–758.

Haas, P.M., 1989. Do regimes matter? Epistemic communities and Mediterranean pollution control. *International Organization*, 43(3), p.377.

Haensch, W., Nowak, E.J., et al., 2006. Silicon CMOS devices beyond scaling. *IBM Journal of Research and Development*, 50(4.5), pp.339–361.

Haensch, W., Nowak, E.J., et al., 2006. Silicon CMOS devices beyond scaling. *IBM Journal of Research and Development*, 50(4.5), pp.339–361.

Ham, R.M., Linden, G. & Appleyard, M.M., 1998. The Evolving Role of Semiconductor Consortia in the United States and Japan. *California Management Review*, 41(1), pp.137–163.

Hart, D.M., 1998. *Forged consensus : science, technology, and economic policy in the United States, 1921-1953*, Princeton University Press.

Heilmeier, G.H., 1984. Microelectronics: End of the beginning or beginning of the end? In *International Electron Devices Meeting*. pp. 2–5.

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

Henderson, R.M. & Clark, K.B., 1990. Architectural Innovation: The Reconfiguration of Existing Product Technologies and the Failure of Established Firms. *Administrative Science Quarterly*, 35(1), p.9.

Hennessy, J.L. & Jouppi, N.P., 1991. Computer technology and architecture: an evolving interaction. *Computer*, 24(9), pp.18–29.

Ho, M.S., Jorgenson, D.W. & Samuels, J.D., 2011. Information Technology and U.S. Productivity Growth: Evidence from a Prototype Industry Production Account. *Journal of Productivity Analysis*, 36(2), pp.159–175.

Hoeneisen, B. & Mead, C.A., 1972. Fundamental limitations in microelectronics—I. MOS technology. *Solid-State Electronics*, 15(7), pp.819–829.

Holbrook, D., 1995. Government Support of the Semiconductor Industry: Diverse Approaches and Information Flows. *Business and Economic History*, 24(2), pp.133–165.

Holdren, J. & Donovan, S., 2016. *National Strategic Computing Initiative Strategic Plan [Available at: https://www.whitehouse.gov/sites/whitehouse.gov/files/images/NSCI%20Strategic%20Plan.pdf]*,

Horowitz, M., 2014. 1.1 Computing's energy problem (and what we can do about it). In *Digest of Technical Papers - IEEE International Solid-State Circuits Conference*. IEEE, pp. 10–14.

HR2401, 1994. *National Defense Authorization Act for Fiscal Year 1994*, House of Representatives.

Hruska, J., 2012. Nvidia deeply unhappy with TSMC, claims 20nm essentially worthless. *ExtremeTech*.

Huang, W. et al., 2011. Scaling with design constraints: Predicting the future of big chips. *IEEE Micro*, 31(4), pp.16–29.

Hughes, T.P., 1983. *Networks of power : electrification in Western society, 1880-1930*, Johns Hopkins University Press.

ICE, 1982. *Status 1981: A Report on the Integrated Circuits Industry*, Phoenix, Arizona.

ICE, 1984. *Status 1983: A Report on the Integrated Circuit Industry*, Phoenix, Arizona.

ICE, 1985. *Status 1984: A Report on the Integrated Circuit Industry*, Phoenix, Arizona.

ICE, 1998. *Status 1997: A Report on the Integrated Circuit Industry*, Phoenix, Arizona.

ICE, 1971. *Status of Integrated Circuits 1970*, Phoenix, Arizona.

Ingram, P., Rao, H. & Silverman, B.S., 2012. History in Strategy Research: what, why, and how? In S. J. Kahl, B. S. Silverman, & M. A. Cusumano, eds. *History and Strategy (Advances in Strategic Management, Volume 29)*. Emerald Group Publishing Limited, pp. 241–273.

Intel Corporation, 2012. Intel and ASML Reach Agreements to Accelerate Key Next-Generation Semiconductor Manufacturing Technologies [Press Release].

IRDS, 2016. *International Roadmap for Devices and Systems*,

ITRS, 2001. *International Technology Roadmap for Semiconductors*, San Jose, CA.

ITRS, 2003. *International Technology Roadmap for Semiconductors*, San Jose, CA.

ITRS, 2005. *International Technology Roadmap for Semiconductors*, San Jose, CA.

ITRS, 2011. *International Technology Roadmap for Semiconductors*, Washington, D.C.

ITRS, 2007. *International Technology Roadmap for Semiconductors*,

ITRS, 2016. *International Technology Roadmap for Semiconductors 2.0*, Washington, D.C.

Jick, T.D., 1979. Mixing Qualitative and Quantitative Methods: Triangulation in Action. *Administrative Science Quarterly*, 24(4), pp.602–611.

Johnson, E., 1965. Physical limitations on frequency and power parameters of transistors. In *IRE International Convention Record*. Institute of Electrical and Electronics Engineers, pp. 27–34.

Jorgenson, D.W., 2001. Information Technology and the U.S. Economy. *American Economic Review*, 91(1), pp.1–32.

Jorgenson, D.W. & Vu, K., 2007. Information Technology and the World Growth Resurgence. *German Economic Review*, 8(2), pp.125–145.

Jouppi, N.P. et al., 2017. In-Datacenter Performance Analysis of a Tensor Processing Unit. *ISCA*.

Kahl, S.J., Silverman, B.S. & Cusumano, M.A., 2012. The Integration of History and Strategy Research. In S. J. Kahl, B. S. Silverman, & M. A. Cusumano, eds. *History and Strategy (Advances in Strategic Management, Volume 29)*. Emerald Group Publishing Limited, pp. ix–xxi.

Kapoor, R., 2013. Persistence of Integration in the Face of Specialization: How Firms Navigated the Winds of Disintegration and Shaped the Architecture of the Semiconductor Industry. *Organization Science*, 24(4), pp.1195–1213.

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

Keyes, R.W., 1975. Physical Limits in Digital Electronics. *Proceedings of the IEEE*, 63(5), pp.740–767.

Keyes, R.W., 1969. Physical problems and limits in computer logic. *IEEE Spectrum*, 6(5), pp.36–45.

Klepper, S., 1996. Entry, Exit, Growth, and Innovation over the Product Life Cycle. *The American Economic Review*, 86, pp.562–583.

Klepper, S., 1997. Industry Life Cycles. *Industrial and Corporate Change*, 6(1), pp.145–182.

Krusius, J.P., 2000. *Si-Based Nanoelectronics. SRC Microscience and Technology Center, Cornell. April 30, 2000. [Report to SRC]*,

Kuhn, T.S., 1962. *The structure of scientific revolutions.*, University of Chicago Press.

Lawler, K., 2012. A New Model for Venture Investment in Semiconductor Start-Ups.

Layton, E.T., 1974. Technology as Knowledge. *Technology and Culture*, 15(1), p.31.

Lecuyer, C., 2005. *Making Silicon Valley: Innovation and the Growth of High Tech, 1930-1970*, Cambridge, Mass.: MIT Press.

Lepadus, M., 2017. Survey: Optimism Grows for EUV. *SemiEngineering*.

Levi, M., 2002. Modeling Complex Historical Processes with Analytical Narratives. In R. Mayntz, ed. *Akteure, Mechanismen, Modelle: Zur Theoriefahigkeit makrosozialer Analyse*.

Levin, R.C. et al., 1987. Appropriating the Returns from Industrial Research and Development. *Brookings Papers on Economic Activity*, 1987(3), p.783.

Levin, R.C., 1982. The Semiconductor Industry. In R. R. Nelson, ed. *Government and technical progress : a cross-industry analysis*. Pergamon Press.

Lim, K., 2009. The many faces of absorptive capacity: spillovers of copper interconnect technology for semiconductor chips. *Industrial and Corporate Change*, 18(6), pp.1249–1284.

Linden, G., Mowery, D.C. & Ham Ziedonis, R., 2000. National Technology Policy in Global Markets: Developing Next-Generation Lithography in the Semiconductor Industry. *Business and Politics*, 2(2), pp.93–113.

Macher, J.T. & Mowery, D.C., 2004. VERTICAL SPECIALIZATION AND INDUSTRY STRUCTURE IN HIGH TECHNOLOGY INDUSTRIES. *Advances in Strategic Management*, 21, pp.317–355.

Macher, J.T., Mowery, D.C. & Hodges, D.A., 1998. Reversal of fortune? The recovery of the U.S. semiconductor industry. *California Management Review*, 41(1), pp.107–136.

Macher, J.T., Mowery, D.C. & Minin, A. Di, 1999. Semiconductors. In D. C. Mowery, ed. *U.S. Industry in 2000: Studies in Competitive Performance*. Washington, D.C.: National Academies Press (US), pp. 245–286.

Macher, J.T., Mowery, D.C. & Di Minin, A., 2007. The``Non-Globalization''of Innovation in the Semiconductor Industry. *California Management Review*, 50(1), p.217.

Mackenzie, D.A., 1990. *Inventing accuracy : a historical sociology of nuclear missile guidance*, MIT Press.

Marshall, A., 2013. Participant Observation of NRI Annual Review on November 8, 2013.

McCray, W.P., 2005. Will small be beautiful? Making policies for our nanotech future. *History and Technology*, 21(2), pp.177–203.

Merrill, R., 1962. Some Society-Wide Research and Development Institutions. In *The Rate and Direction of Inventive Activity: Economic and Social Factors*. Princeton University Press, pp. 409–440.

Mody, C.C.M., 2016. *The long arm of Moore's law : microelectronics and American science*, MIT Press, Cambridge, MA.

Mollick, E., 2006. Establishing Moore's Law. *IEEE Annals of the History of Computing*, 28(3), pp.62–75.

Moore, G.E., 1965. Cramming More Components onto Integrated Circuits. *Electronics*, 38(8), pp.114–117.

Moore, G.E., 1975. Progress in digital integrated electronics. In *1975 International Electron Devices Meeting*. pp. 11–13.

Morton, J.A., 1965. From physics to function. *IEEE spectrum*, 2(9), pp.62–66.

Mowery, D.C., 2011. Nanotechnology and the US national innovation system: continuity and change. *The Journal of Technology Transfer*, 36(6), pp.697–711.

Mowery, D.C., 2009. Plus ca change: Industrial R&D in the "third industrial revolution." *Industrial and Corporate Change*, 18(1), pp.1–50.

Mowery, D.C. & Rosenberg, N., 1991. *Technology and the Pursuit of Economic Growth*, Cambridge University Press.

Mowery, D.C. & Teece, D.J., 1996. Strategic alliances and industrial research. In R. Rosenbloom & W. J. Spencer, eds. Harvard Business Review Press.

Murmann, J.P. & Frenken, K., 2006. Toward a systematic framework for research on dominant designs, technological innovations, and industrial change. *Research Policy*, 35(7), pp.925–952.

NACS, 1992a. *A national strategy for semiconductors: an agenda for the President, the Congress, and the industry*, Arlington, Va. : Washington, DC: National Advisory Committee on Semiconductors.

NACS, 1992b. Attaining preeminence in semiconductors: third annual report to the President and the Congress. , p.2 v.

NACS, 1991a. *Micro Tech 2000 Workshop Report*,

NACS, 1991b. *Toward a national semiconductor strategy: second annual report of the National Advisory Committee on Semiconductors.*, Arlington, Va. : [Washington, D.C.: National Advisory Committee on Semiconductors.

Natarajan, S. et al., 2014. A 14nm logic technology featuring 2nd-generation FinFET, air-gapped interconnects, self-aligned double patterning and a 0.0588 μm2 SRAM cell size. In *2014 IEEE International Electron Devices Meeting*. IEEE, p. 3.7.1-3.7.3.

National Academies of Sciences, Engineering, and M., 2016. *Triennial Review of the National Nanotechnology Initiative*, Washington, DC: The National Academies Press.

National Center for Science and Engineering Statistics, 2017. Survey of Federal Funds for Research and Development - NCSES - US National Science Foundation (NSF).

National Research Council, 1999. *Funding a Revolution: Government Support for Computing Research*, Washington, D.C.: National Academies Press.

National Research Council, 2003. *Securing the Future:Regional and National Programs to Support the Semiconductor Industry* C. W. Wessner, ed., Washington, D.C.: The National Academies Press.

National Research Council, 2011. *The Future of Computing Performance*, Washington, D.C.: National Academies Press.

Nelson, R.R., 1962. Introduction to "The Rate and Direction of Inventive Activity: Economic and Social Factors." In *The Rate and Direction of Inventive Activity: Economic and Social Factors*. Princeton University Press, pp. 1–16.

Nelson, R.R., 1993. *National Innovation Systems: A Comparative Analysis*, Rochester, NY.

Nelson, R.R., 2004. The market economy, and the scientific commons. *Research Policy*, 33(3), pp.455–471.

Nelson, R.R., 1959. The Simple Economics of Basic Scientific Research. *Journal of Political Economy*, 67.

Nelson, R.R., 1961. Uncertainty, Learning, and the Economics of Parallel Research and Development Efforts. *The Review of Economics and Statistics*, 43(4), p.351.

Nelson, R.R. & Winter, S.G., 1982. *An evolutionary theory of economic change*, Belknap Press of Harvard University Press.

Nelson, R.R. & Winter, S.G., 1977. In search of useful theory of innovation. *Research Policy*, 6(1), pp.36–76.

Nikonov, D.E., 2013. Interview by author at NRI Annual Review on October 22, 2013.

Nikonov, D.E. & Young, I.A., 2015. Benchmarking of Beyond-CMOS Exploratory Devices for Logic Integrated Circuits. *IEEE Journal on Exploratory Solid-State Computational Devices and Circuits*, 1, pp.3–11.

Nikonov, D.E. & Young, I.A., 2013. Overview of Beyond-CMOS Devices and a Uniform Methodology for Their Benchmarking. *Proceedings of the IEEE*, 101(12), pp.2498–2533.

NIST, 2007. NIST CONSORTIUM/CONSORTIA FOR POST-COMPLEMENTARY METAL OXIDE SEMICONDUCTOR (CMOS) NANOELECTRONICS RESEARCH PROGRAM; AVAILABILITY OF FUNDS. Federal Register Volume 72, Number 86 (May 4, 2007).

NSCI, 2016. *National Strategic Computing Initiative (NSCI) Strategic Plan*,

NSET, 2010. *NSI White Paper: Nanoelectronics for 2020 and Beyond*, Washington, D.C.

NSF, 2004. *Nanoscale Science and Engineering (NSE) Program Solicitation for FY 2005*,

NSF, 2017. National Science Foundation Awards Database. http://www.nsf.gov/awardsearch/advancedSearch.jsp Accessed 3/1/2017.

NTRS, 1994. *National Technology Roadmap for Semiconductors*,

NTRS, 1997. National Tecnology Roadmap for Semiconductors.

Ouchi, W.G., 1984. Political and Economic Teamwork: The Development of the Microelectronics Industry of Japan. *California Management Review*, 26(4), p.8.

Packan, P.A., 1999. Pushing the Limits. *Science*, 285(5436), pp.2079–2081.

Pavitt, K., 1984. Sectoral patterns of technical change: Towards a taxonomy and a theory. *Research Policy*, 13(6), pp.343–373.

PCAST, 2017. *REPORT TO THE PRESIDENT Ensuring Long-Term U.S. Leadership in Semiconductors*, Washington, D.C.

Platzer, M. & Sargent Jr., J., 2016. *U.S. Semiconductor Manufacturing: Industry Trends, Global Competition, Federal Policy (CRS Report: R44544)*, Washington, D.C.

Polanyi, M., 1958. *Personal knowledge. Towards a Post-Critical Philosophy.*, University of Chicago Press.

Powell, W.W. & Grodal, S., 2005. Networks of innovators. In J. Fagerberg, D. C. Mowery, & R. R. Nelson, eds. *The Oxford Handbook of Innovation*. New York: Oxford University Press, pp. 56–85.

Powell, W.W., Koput, K.W. & Smith-Doerr, L., 1996. Interorganizational Collaboration and the Locus of Innovation: Networks of Learning in Biotechnology. *Administrative Science Quarterly*, 41(1), pp.116–145.

Randall, J.N., Reed, M.A. & Frazier, G.A., 1989. Nanoelectronics: Fanciful physics or real devices? *Journal of Vacuum Science & Technology B: Microelectronics and Nanometer Structures*, 7(6), p.1398.

RDSS, 2017. *Research and Development Support System: DARPA Basic Research Budget,2001-2016 [Accessed 3/7/2017]*,

Riordan, M. & Hoddeson, L., 1997. *Crystal fire: the invention of the transistor and the birth of the information age*, WW Norton & Company.

Roco, M.C., 2011. The Long View of Nanotechnology Development: The National Nanotechnology Initiative at 10 Years. In Science Policy Reports. Springer Netherlands, pp. 1–28.

Rosenberg, N., 1982. *Inside the black box : technology and economics*, Cambridge University Press.

Rosenberg, N., 1976. On Technological Expectations. *Economic Journal*, 86(343), pp.523–35.

Rosenberg, N., 1969. The Direction of Technological Change: Inducement Mechanisms and Focusing Devices. *Economic Development and Cultural Change*, 18(1), pp.1–24.

Rosenberg, N. & Nelson, R.R., 1994. American universities and technical advance in industry. *Research Policy*, 23(3), pp.323–348.

Rosenkopf, L. & Almeida, P., 2003. Overcoming Local Search Through Alliances and Mobility. *Management Science*, 49(6), pp.751–766.

Rosenkopf, L. & Tushman, M.L., 1998. The Coevolution of Community Networks and Technology: Lessons from the Flight Simulation Industry. *Industrial and Corporate Change*, 7(2), pp.311–346.

Sahal, D., 1985. Technological guideposts and innovation avenues. *Research Policy*, 14(2), pp.61–82.

Sakakibara, K., 1983. *From imitation to innovation : the very large scale integrated (VLSI) semiconductor project in Japan*, Cambridge, Mass. : Massachusetts Institute of Technology.

Sakakibara, K., 1993. R&D cooperation among competitors: A case study of the VLSI semiconductor research project in Japan. *Journal of Engineering and Technology Management*, 10(4), pp.393–407.

Schaller, R.R., 2004. *Technological innovation in the semiconductor industry: A case study of the International Technology Roadmap for Semiconductors (ITRS)*. George Mason University.

Scherer, F.M., 2011. *Parallel R&D Paths Revisited*, Cambridge, MA.

Schmookler, J., 1962. Changes in Industry and in the State of Knowledge as Determinants of Industrial Invention. In *The Rate and Direction of Inventive Activity: Economic and Social Factors*. Princeton University Press, pp. 195–232.

Schumpeter, J.A., 1942. *Capitalism, socialism, and democracy*, New York: Harper and Brothers.

Semiconductor Industry Association, 1981. *1980-1981 Yearbook and Directory.*, Cupertino, CA.

SIA, 2004. NRI Scope and Objective [Presentation to SIA TSC June 9, 2004].

Solow, R.M., 1957. Technical Change and the Aggregate Production Function. *The Review of Economics and Statistics*, 39(3), p.312.

Spencer, W.J. & Grindley, P., 1993. SEMATECH after five years: high-technology consortia and US competitiveness. *California Management Review*, 35, pp.9–9.

Spencer, W.J. & Seidel, T.E., 2004. International technology roadmaps: The US semiconductor experience. In *Productivity and Cyclicality in Semiconductors:: Trends, Implications, and Questions--Report of a Symposium*. p. 135.

Sperling, E., 2017. Is 7nm The Last Major Node? *SemiEngineering*.

SRC, 1983. *1983 Annual Report*, Research Triangle Park, NC.

SRC, 2013. NRI Mission and Description. Available at: https://www.src.org/program/nri/about/mission/ [Accessed October 1, 2016].

SRC, 2003. *Research Needs for Novel Devices.*, Research Triangle Park, NC.

SRC, 2005. Semiconductor Industry Announces New Nanoelectronic Research Grants to U.S. Universities [Press Release]. Available at: https://www.src.org/newsroom/press-release/2005/27/ [Accessed September 10, 2013].

SRC, 2000. Solicitation for Concepts for Research: Novel Devices for Information Processing. Available at: https://www.src.org/compete/archive/grc/2000-novel-device/ [Accessed October 20, 2014].

SRC, 2006a. SRC, Texas Instruments and the State of Texas Launch World-Class Nanoelectronics Research Initiative. Available at: https://www.src.org/newsroom/press-release/2006/36/ [Accessed September 10, 2013].

SRC, 2012. *SRC 2012 Annual Report. 30 Years: Proof of Effectiveness*, Research Triangle Park, NC.

SRC, 1998. *SRC Annual Report*, Research Triangle Park, NC.

SRC, 2006b. Texas Instruments and the State of Texas Launch World-Class Nanoelectronics Research Initiative [Press Release]. Available at: http://www.src.org/newsroom/press-release/2006/36/ [Accessed September 10, 2013].

SRC, 2008. World-Class Nanoelectronics Research Center Launched by Semiconductor

Research Corporation (SRC), Nanoelectronics Research Initiative (NRI) [Press Release]. Available at: https://www.src.org/newsroom/press-release/2008/49/ [Accessed September 10, 2013].

Standard & Poor's, 2013. Complete Financial Statements Advanced Micro Devices Inc, Intel Corportation, International Business Machines Corp., Micron Technology, Inc., Texas Instruments Inc. from Compustat Database. Retrieved May 14, 2013.

STC, 1996. *Semiconductor Technology Council: First Annual Report*, Washington, D.C.

Sumney, L.W. & Burger, R.M., 1986. The Semiconductor Research Corporation and University Research in Integrated Circuits. *IEEE Transactions on Education*, E-29(2), pp.61–68.

Sydow, J. et al., 2012. Organizing R&D Consortia for Path Creation and Extension: The Case of Semiconductor Manufacturing Technologies. *Organization Studies*, 33(7), pp.907–936.

Theis, T., 2013. Interview by author at IBM TJ Watson Lab on July 19, 2013.

Theis, T.N. & Solomon, P.M., 2010. It's Time to Reinvent the Transistor! *Science*, 327, p.1600.

Thompson, N., 2012. *Firm Software Parallelism: Building a measure of how firms will be impacted by the changeover to multicore chips*. University of California, Berkeley.

Thompson, N., 2017. The Economic Impact of Moore's Law: Evidence from When it Faltered. *SSRN Electronic Journal*.

Tilton, J.E., 1971. *International Diffusion of Technology: The Case of Semiconductors*, Brookings Institution Press.

Tushman, M.L. & Anderson, P., 1986. Technological Discontinuities and Organizational Environments. *Administrative Science Quarterly*, 31(3), p.439.

Tushman, M.L. & Rosenkopf, L., 1992. Organizational determinants of technological-change-toward a sociology of technological evolution. *Research in Organizational Behavior*, 14, pp.311–347.

Utterback, J.M., 1994. *Mastering the dynamics of innovation: how companies can seize opportunities in the face of technological change*, Boston (Mass.): Harvard Business School Press.

Utterback, J.M. & Abernathy, W.J., 1975. A dynamic model of process and product innovation. *Omega*, 3(6), pp.639–656.

Utterback, J.M. & Suárez, F.F., 1993. Innovation, competition, and industry structure. *Research Policy*, 22(1), pp.1–21.

Vincenti, W.G., 1990. *What engineers know and how they know it : analytical studies from aeronautical history*, Johns Hopkins University Press.

Wallmark, J. & Marcus, S., 1962. Minimum Size and Maximum Packing Density of Nonredundant Semiconductor Devices. *Proceedings of the IRE*, 50(3), pp.286–298.

Welser, J.J., 2013. Interview by Author at IBM Research Almaden on December 19, 2013.

Welser, J.J. et al., 2008. The quest for the next information processing technology. *Journal of Nanoparticle Research*, 10(1), pp.1–10.

Winter, S.G., 2005. *Developing Evolutionary Theory for Economics and Management*,

WTEC, 1998. *Nanostructure Science and Technology: A Worldwide Study*, Arlington, Va.

Yoffie, D.B., 1988. How An Industry Builds Political Advantage. *Harvard Business Review*.

Zhirnov, V.V. et al., 2003. Limits to binary logic switch scaling-a gedanken model. *Proceedings of the IEEE*, 9(11), pp.1934–1939.

Appendix A: Data Sources from the SRC

Data Type	Data Source	Data Details	Data Availability	Size (N)
SRC Inputs & Institutions	SRC Research Contracts	Database of all SRC research contracts with universities including every associated researcher (professors, students, company liaisons)	1982-2013	3500 research contracts
SRC Institutions	SRC Corporate Members	List of SRC corporate members with information on types of membership	1982-2013	212 organizations
SRC Individuals	SRC Student Database	Database of all students currently or formerly (student alumni) funded by an SRC contract	1982-2013	10,000 students
SRC Individuals	SRC Technical Advisory Board (TAB) Database	Time series of all company representatives on SRC technical, advisory, and executive boards	1982-2013	4700 representatives
NRI Institutions & Individuals	NRI (SRC program started in 2006)	Full list of attendees of NRI conferences, including member company representatives	2006-2013	700 attendees
SRC-Specific Outputs	SRC's Internal Documentation of its Outputs	SRC list of all patent approvals and filings associated with its contracts	1982-2013	755 patent filings
SRC-Specific Outputs	SRC's Internal Documentation of its Outputs	SRC list of all reports and working papers associated with its contracts	1982-2013	14000 papers