HARVARD | BUSINESS | SCHOOL

Explaining the Vertical-to-Horizontal Transition in the Computer Industry

Carliss Y. Baldwin

Working Paper 17-084

Explaining the Vertical-to-Horizontal Transition in the Computer Industry

Carliss Y. Baldwin
Harvard Business School

Working Paper 17-084

Explaining the Vertical-to-Horizontal Transition in the Computer Industry

Excerpts from:
Design Rules Volume 2—
How Technology Shapes Organizations

Carliss Y. Baldwin

March, 2017

Abstract

This paper seeks to explain the technological forces that led to the rise of vertically integrated corporations in the late 19th Century and the opposing forces that led to a vertical-to-horizontal transition in the computer industry one hundred years later. I first model the technology of step processes with bottlenecks and show how this technology rewards vertical integration, a hierarchical organization, and the use of direct authority. These properties in turn became the organizational hallmarks of so-called "modern" corporations. I then model platform systems, showing that, in contrast to step processes, this technology rewards the multiplication of options, increasing risk, and modularity. Moreover, given a modular architecture, a platform system can be *open*, with different components supplied by separate firms with no loss of interoperability or efficiency. Openness multiplies options and expands diversity, thus increasing the platform system's value. The last two decades of the 20th Century saw the rise of three distinct types of open platforms in the computer industry: (1) "forward open" platforms with downstream complementors; (2) "backward open" modular supply networks; and (3) "open exchange" platforms designed to facilitate transactions and other forms of social interaction. Whereas in 1980, vertically integrated firms dominated the industry, by 2000, the "verticals" had essentially disappeared. The largest firms in the industry in 2000 were sponsors and participants in open platform systems. I argue that the vertical-to-horizontal transition in the computer industry was an organizational response to a fundamental change in economic rewards to the technologies of rationalized step processes vs. open platform systems.

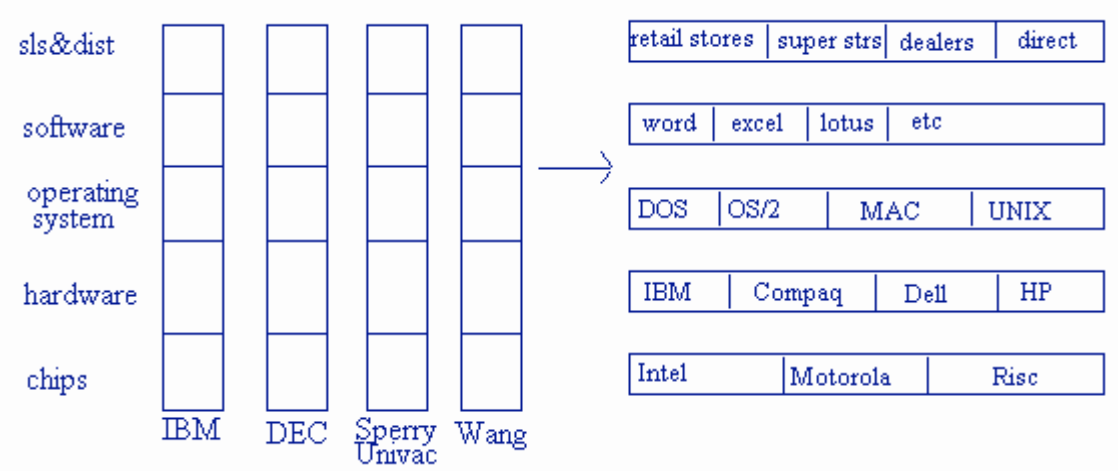
1 Introduction

This paper seeks to explain the technological forces that led to the rise of vertically integrated corporations in the late 19th Century and the opposing forces that led to the vertical-to-horizontal transition in the computer and other industries one hundred years later. To construct this argument, I have taken excerpts from my book manuscript: *Design Rules, Volume 2: How Technology Shapes Organizations*. This work is ongoing and comments on any aspect are most welcome.

1.1 A Puzzle: The Vertical-to-Horizontal Transition in the Computer Industry

In 1995, Andy Grove described a vertical-to-horizontal transition in the structure of the computer industry. In a now-famous picture (Figure 1-1), he described the transformation of that industry from a set of vertically integrated "silos", e.g., IBM, DEC, Sperry Univac and Wang, to a large number of firms organized in functional "layers," e.g., the chip layer, the computer layer, the operating system, application software, and sales and distribution layers.

Figure 1-1
The Vertical-to-Horizontal Transition in the Computer Industry



Source: Adapted from Grove (1996) *Only the Paranoid Survive*, p. 44.

Grove did not know what caused this transition, although he felt it was spurred by increasingly cheap integrated circuits (Moore's Law) plus the possibilities of modular recombinations:

> A consumer could pick a chip from the horizontal chip bar, pick a consumer manufacturer from the computer bar, choose an operating system … grab one of the several ready-to-use applications off the shelf,… and take the collection of these things home. … He might have trouble making them work,… but for $2000 he had just bought a computer system.[1]

But though the causes of the transition were unclear, Grove could see that the consequences for the industry were profound:

> Going into the eighties, the old computer companies were strong, growing and vital. ... But by the end of the eighties, many large vertical computer companies were in the midst of layoffs and restructuring…
>
> [A]t the same time, the new order provided an opportunity for a number of new entries to shoot into preeminence. ...[For example] Compaq … understood the dynamics of the new industry and prospered by tailoring their business model to it.[2]

Today, we would call Grove's horizontal industry structure a *business ecosystem* organized around *platforms* which in turn are built on *modular technical architectures*.[3] In this type of industry, most firms make *modules* that are in turn parts of larger systems, as Grove's horizontal figure indicates. Ecosystems based on modular architectures can include hundreds or even thousands of firms producing different functional components that are combined into systems by users or specialized systems integrators.[4]

The ecosystem form of industry structure emerged in the computer industry between 1985 and 1995. The consequences of the transition were indeed vast—in terms of value created for consumers, value created and destroyed for investors, and turnover in entry and exit. Today, other industries, e.g., telecommunications and pharmaceuticals, may be undergoing similar vertical-to-horizontal transitions.

[1] Grove (1996) pp. 41-42.

[2] Ibid. p. 45.

[3] On ecosystems, see for example, Moore (1996); Iansiti and Levien (2004); Adner and Kapoor (2010); Jacobides, Cennamo and Gawer (2017). On platforms, see Bresnahan and Greenstein (1999); Gawer and Cusumano (2002); Parker, Van Alstyne and Choudary (2016). On modular technical architectures, see Langlois and Robertson (1990); Sanchez and Mahony (1996); Garud and Kumaraswamy (1995); Baldwin and Clark (2000); Schilling (2000); and Sturgeon (2002).

[4] Brusoni, Prencipe and Pavitt (2001); Prencipe, Davies and Hobday (2003); Berger (2005).

But we understand very little about what causes such changes to take place.

In fact, we lack understanding on two levels. The vertical form of industry structure is itself a relatively new phenomenon. Vertically integrated firms came to dominate many industries in the United States in the late 19th and early 20th Centuries.[5] During the 20th Century, such "giant enterprises" expanded into new product markets and new geographies. They were run by salaried managers and organized as bureaucratic hierarchies. For better or worse, in the minds of most observers, such firms epitomized "big business" and "high technology."[6]

2 The Technology of Flow Rationalization

From 1750 to 1920 approximately, the movement of industrial organization was towards higher levels of centralization supporting ever-higher levels of output. First, production tasks came to be located in factories, which replaced the older "putting out" of household production. In factories, technical steps were subdivided and human strength augmented by powered machinery. Even though Adam Smith praised the productivity gains of a simple division of labor, the need to be near a central power source seems to have been the decisive factor behind the rise of factories in England in the 18th Century.[7]

In the decades following 1750, machines got better, and sources of power, including water and steam engines, became more efficient. Larger and more powerful machines were inserted into production processes that were divided into finer steps. Around 1850, the baton of industrial innovation passed from Great Britain to the United States.[8] The American railroad system made the Midwest and far West accessible to Eastern producers. Transportation costs and shipment times dropped precipitously so that

[5] Chandler (1962; 1977; 1990)

[6] Schumpeter (1947); Kaplan (1964); Galbraith (1967); Servan-Schrieber (1969), Rosenberg and Birdzell (1986); Hounshell (1988); Drucker (1993); L. Marx (1994); Baumol (2002); Landes (2003); Fukuyama (1995).

[7] Landes (1998) p. 209.

[8] Hounshell (1984) pp. 17-25.

many local markets could now be served by distant companies.[9] The so-called American system of manufacturing, which was based on very fine divisions of labor and specialized machinery, permitted single factories to produce goods in volumes never seen before.

However as the tasks became more subdivided and the intermediate steps more numerous, production systems, like that at the Singer Sewing Machine Co., spun out of control.

> The problem was a gradual breakdown of the integration of work flow at the lower levels of the company and a concordant deterioration in the ability of top executives to control work lower in the company hierarchy.[10]

The response was a movement towards "systematic management" aimed at rationalizing production within factories.[11] The systematizers, including Slater Lewis, Henry Metcalf, Alexander Church, H.M. Norris, and John Tregoing, invented production control systems, inventory control systems, and cost accounting systems and implemented them at a number of firms.[12] Frederick W. Taylor extended their work by incorporating detailed time studies and advocating a "differential piece rate" form of compensation. He became a famous (and in some circles infamous) advocate of what he called "scientific management." (Taylor was the most visible and controversial proponent of efficiency and flow rationalization, but the systematizing movement started well before his career began. Most of the important innovations were made by others. In fact, despite Taylor's prestige, his system of shop management was seldom implemented.)[13]

These organizational innovations aimed at controlling and coordinating the flow of production through a factory increased output, but also took away the workers' autonomy. Rather than the foreman or

---

[9] Chandler (1977); Fields (2004).

[10] Litterer (1963) p. 373.

[11] Kendall, H. P. (1911) "Types of Management: Unsystematized, Systematized, and Scientific," in *Dartmouth College Conferences, Addresses and Discussions of the Conference on Scientific Management Held October* (Vol. 12, No. 73, p. 14).

[12] Litterer (1963) p. 370.

[13] D. Nelson (1974); Chandler (1977) pp. 272-277. Misa (1995; p. 208) suggests that many firms had an ulterior motive for hiring Taylor as a consultant. "[T]aking steps to install Taylor's shop management scheme was the (only) means of acquiring knowledge of the invaluable metal-cutting experiments, including the latest on high-speed-tool steel." See Chapter 7 on the impact of high-speed steel on the design of machine tools. Even after the Taylor-White patents were voided, Taylor still controlled reams of experimental data on metal-cutting techniques.

worker deciding what to make and what supplies to use, control and scheduling functions were performed by specialized staff, including stock clerks (to control inventory), production control clerks (to keep track of orders) and time keepers (to measure work flow through various tasks).[14] The methods and principles developed by the systematizers allowed complex factory systems to operate at ever higher levels of output (scale) without collapsing into chaos.

For example, in oil, the throughput of the largest refineries increased from 500 to 6500 barrels per day between 1866 and 1879, and increases continued through 1900.[15] In steel, the production of large blast furnaces increased from about 6000 tons per year in the 1860s to over 100,000 tons a year in the late 1890s.[16] Rolling mills exhibited an even higher rate of growth in throughput: in 1850 a typical mill might produce 3000 tons a year, while in 1900 a large rolling mill's output was 3000 tons *a day*, or approximately 900,000 tons a year.[17] A single Bonsack machine could roll as many cigarettes as 50 skilled workers at a fraction of the cost.[18] The output of Singer sewing machines was less that 1000 per year in 1856; by 1880, two factories produced 500,000 machines per year.[19]

A flow of tasks and transfers is a necessary corollary of a division of labor. What was previously undivided work performed by one person becomes a set of tasks performed by different people and a series of transfers between them.[20] Below I describe why it is necessary to rationalize flow processes, and then explore how the technologies of flow production shaped the boundaries and internal structure of organizations.

---

[14] Litterer (1963) *op. cit.* p. 387.

[15] Chandler (1985) "The Standard Oil Company—Combination, Consolidation, and Integration," in *The Coming of Managerial Capitalism: A Casebook on the History of American Economic Institutions*, (A.D Chandler, Jr. & R.S. Tedlow, eds.) Homewood IL: Richard D. Irwin, pp. 343-371.

[16] Temin (1964) p. 159.

[17] *Ibid.* p. 165; Popplewell (1906) p.103.

[18] http://www.learnnc.org/lp/editions/nchist-newsouth/4705 (accessed June 13, 2016).

[19] Hounshell (1984) p. 89-123.

[20] Smith, A. (1994) Book 1, Chapter 1.

2.1 A Model of Production based on Flow

Let me model a flow production process as a series of steps that begins by taking in raw materials, fabricates intermediate components, combines components into a finished product, and then transports and sells the product to the final customer. The steps take place in a sequence, although not always a strict sequence. (A strict sequence exists in many "continuous flow" processes, including paper-making, the production of iron and steel, and textile manufacturing. However, before and after the strictly sequential steps, other steps can take place in parallel. For example, intermediate components can be produced in different places in a factory and then assembled. Ingots of metal may be formed in the same furnace, but then sent to different rolling mills to be fabricated into finished products. In a "job shop," the sequence of steps may vary from job to job, depending on the specifications of the order. Steps are essential to the model, sequence is not.)

When a single factory makes many products, scheduling the flow of production is very difficult. When many steps are involved, it becomes impossible to keep all machinery and workers fully utilized at all times. Thus it is not surprising that, as the use of expensive machinery increased, many firms reduced the breadth of their product lines.

> At one time a metal working factory would be willing to make pumps, steam engines, farm implements, tools, locomotives, in brief, just about anything in metal their craftsmen could handle. By the end of the Civil War a number of specialized manufacturers emerged who made just pumps, or locomotives or machine tools.[21]

The critical property of a series of steps aimed at making a particular product is that all steps must take place in strict proportion. Let us think of the productive steps as a set of functional complements:

$$s_1 \square s_2 \square \ldots s_i \square s_j \ldots \square s_{N-1} \square s_N = S \qquad (1)$$

Here s_i denotes a single step in the production process and S the finished good, which might be a sewing machine. The symbol $\square$ signifies that the steps are strong complements, i.e., if any step is left out the entire effort fails. Within each step, a particular technical procedure is carried out and the intermediate

[21] Litterer (1961) p. 467.

good is then passed to another step.[22] Each step is performed by a worker in conjunction with appropriate materials and machinery.

The steps are tied together by more than functional complementarity. Each has a certain *capacity*, that is, a maximum number of units that can be processed per unit of time. Because all steps are necessary to make the final good, the *production capacity of the entire system* in a given time interval equals the minimum of the capacities of the separate steps:

$$Q_{\min} = \min(q_1,...,q_i,q_j,...,q_N) \quad . \tag{2}$$

Here $Q_{\min}$ denotes the capacity of the system; and q_i denotes the capacity of step s_i. In general, individual q_i' s are stochastic, that is, the output of each step in each time interval may vary because of unknown causes. Such variation can be addressed through systematic diagnosis and problem-solving.

The step with the least capacity (in any interval) is known as the *production bottleneck*:

$$\text{Production Bottleneck} \equiv \text{step such that } q_B = \min(q_1,...,q_N) \ . \tag{3}$$

As with all bottlenecks, production bottlenecks are important targets of managerial attention and investment.

2.2 Properties of Stochastic Step Processes

Two important properties of stochastic multi-step processes can be shown to hold for any set of underlying distributions. They are:

Proposition S-1. *In the absence of rationalization, expected system capacity decreases with the number of steps in the process*. In other words, adding steps by subdividing work without attending to bottlenecks is likely to make overall performance worse.[23]

---

[22] Even steps that take place in parallel are functional complements. Thus the index may not correspond to the timing or order of steps.

[23] **Proof of Proposition S-1.** Consider one realization of a process with N steps. The realization results in a capacity for the system as a whole, $\hat{Q}$, that is the minimum of the realizations of the N steps:

$$\hat{Q} = \min(\hat{q}_1,...,\hat{q}_N) \ .$$

Now consider adding a step to the process. The new step has a cumulative distribution function $F_{N+1}(q_{N+1})$. This function does

Proposition S-2. *In the absence of rationalization, expected system capacity decreases with the random variability of any step.* Thus adding random variation to any step is also likely to make overall performance worse.[24]

2.3 Implications of Proposition S-1

Proposition S-1 implies that the division of labor is a two-edged sword. On the one hand, the narrowing of tasks combined with special-purpose, powered machinery can greatly increase the capacity of an individual step. However, the process as a whole is hostage to the least-efficient step—the production bottleneck. Especially in systems using novel technology and/or experiencing rapid growth, adding more steps to the process has the potential to decrease the capacity of the entire system in the short run.

The solution to this conundrum, of course, is not to take the capacity of any step as a given. Instead managers must proactively seek to identify production bottlenecks and increase their capacity. This means first studying the process from start to finish. In the 19th and early 20th Centuries the problem was addressed by men armed with stopwatches, clipboards, and slide rules. Firms began to hire special timekeepers, process engineers, and ultimately planners and schedulers to observe the workers and

---

not have to be known to the analyst. Let the support of F_{N+1} be ($q_{\min}$, $q_{\max}$). If $q_{\min} < \hat{Q}$, then adding step *N+1* diminishes the capacity of the system with probability $F_{N+1}(\hat{Q})$, which is greater than zero. If $\hat{Q} \le q_{\min}$, then adding the step leaves system capacity unchanged. Thus adding a step weakly decreases the expected capacity of the system as a whole. *QED.*

[24] **Proof of Proposition S-2.** Consistent with Rothschild and Stiglitz (1970), I define increasing variability (risk) as the addition of a mean preserving spread to a given probability distribution. Consider again a process of *N* steps that has a realized capacity of $\hat{Q} = \min(\hat{q}_1, ..., \hat{q}_N)$. The impact of the *N+1* step on the capacity of the system is:

$$P(\tilde{q}_{N+1}) = \begin{cases} \hat{q}_{N+1} - \hat{Q} & \text{if } \hat{q}_{N+1} < \hat{Q} \\ 0 & \text{otherwise} \end{cases}.$$

$P(\cdot)$ is a concave function, thus, as demonstrated by Rothschild and Stiglitz:

$$EP(Y) \le EP(X)$$

if $Y = X +$ mean preserving spread. Therefore increasing the variability of the $N+1^{st}$ step, weakly reduces the expected value of system capacity. (The reduction is strong if the lower bound of the support of *Y* is less than $\hat{Q}$. This result holds (1) for any focal step in the process, and (2) for any realization of $\hat{Q}$ for the *N* steps that are not the focal step. *QED.*

machines, analyze the data, and implement flow-enhancing changes in the technical design of the systems.[25]

These men (for most were men) inserted a new layer of specialized workers between the top managers of a factory and the foremen and workers who handled material and machines. The new layer of staff dealt with information—orders, schedules, inventory, plans— which was used to eliminate bottlenecks and rationalize flow within and beyond the factory. The hiring of these individuals signaled the emergence of a multi-level, multi-function managerial hierarchy. These clerks were the precursors of a new class of middle managers in what became large corporate bureaucracies.[26]

2.4 Implications of Proposition S-2

According to Proposition 1, a random negative draw in any step may turn that step into a bottleneck. Proposition 2 then states that the *wider* the potential variation, the more damage a random bottleneck can do to the performance of system as a the whole. Uncontrolled variation is "the enemy" in a multi-step production process subject to capacity contraints.

It follows that there is real value to *controlling* each step to reduce its intrinsic variation.[27] In effect, the technical system creates an environment in which extreme risk aversion pays. This is the essential insight behind the so-called "six sigma" approach to process improvement: the process performs as predicted 99.99966% of the time.[28]

If particular steps cannot be controlled beyond certain limits, then there must be buffers between them. If the individual step-capacities in each small time interval are independent, then the variance of the

[25] Not everyone saw value in these procedures. The owners of smaller shops, in particular, perceived the methods to be "theoretical and highly impractical," leading only to "extra clerks." One foreman said of the systematizers: "they had every man in the place running around with a pencil over his ear, and we didn't get the work done." (Shenhav, 1995, p. 565).

[26] Litterer (1963) pp. 68-69; Chandler (1977) pp. 273-283.

[27] Jaikumar and Bohn (1986) have proposed a dynamic theory of production in which knowledge proceeds by stages from mere recognition of "good" vs. "bad" processes, to recognition of attributes affecting performance, to partial control of attributes, to recognition of contingencies, to partial control of contingencies. The way to improve such systems is to drive out all uncertainty, but such complete control is seldom, if ever, achieved in practice.

[28] Harry and Schroeder (2005)

sequence will decrease as the time interval grows longer. In this fashion, buffers can absorb the variability of individual steps and the law of large numbers can work to make the throughput of the system more consistent and predictable. Thus buffering is a way to reduce effective step variability and increase the overall capacity of the system.

However, buffering comes at a cost. First, there is the direct cost of the inventory itself. Second, the use of buffers makes it unnecessary to study each step in detail to reduce its variation through better design of the actual work flow. If one can reduce step variation directly, then buffering inventories can be eliminated. This is the essential insight behind Toyota's identification of buffers as source of "waste" (muda) in a production system.

In a multi-step flow process, technology operating through the "min" function ties the steps together in a particular way so that any step can constrain the whole system. Interdependence among steps is what makes variability costly for the process as a whole. To see this, suppose each step made an completely independent contribution to the value of the whole. Then if one step had poor throughput in a given interval, the other steps would not be affected. Over time, each step would contribute an average amount to the process as a whole, and the output of the whole would be the sum of these averages.

3 The Elimination of Bottlenecks Requires Vertical Integration

The presence of bottlenecks scattered through the steps of a process creates the need for centralized coordination and integrated decision-making within the boundaries of a single firm. In the absence of bottlenecks, different parts of a step process can be carried out with little or no loss of value by two or more separate firms. To see this, let us imagine a production process with labor-saving capital and a growing market, *but no bottlenecks*. Consider two firms, Upstream and Downstream. Upstream is responsible for production and can make labor-saving capital investments. Downstream purchases Upstream's output and is responsible for distribution and marketing: it can make investments that increase the size of the market. By assumption, there are no bottlenecks: either type of firm can increase

its capacity without limit.

I assume that the Upstream and Downstream have structured their contracts to avoid the problem of "double marginalization." This can be done via several types of revenue-sharing arrangements: for example, Downstream might receive a commission on sales of Upstream products, or it might apply a consistent percentage markup, or it might pay Upstream a percentage royalty. I also assume that Upstream and Downstream's assets are not co-specialized: each has other potential trading partners thus is not vulnerable to holdup by the other. However, through product differentiation, the Upstream-Downstream combination exercises market power, and chooses price and quantity to maximize their profits. (With revenue-sharing, their choices will be consistent.)

Now suppose Downstream has the opportunity to invest in advertising to increase the size of its market. In the absence of bottlenecks, it will increase the quantity of goods ordered from its Upstream supplier. Thus the Upstream firm will benefit from Downstream's market-expanding investment (an externality). The higher quantities ordered will then increase the value to Upstream of cost-saving capital investment. Once the investment has been made, lower variable costs will lead Upstream to reduce the price of its intermediate good. Thus Downstream will benefit from Upstream's cost-reducing investment (another externality). It will then have incentive to invest again to expand the market.

In effect, Upstream and Downstream firms *are in a symbiotic relationship, but they can coordinate their actions in the market, via prices charged and quantities ordered*. Even though their choices display supermodular complementarity,[29] market signals are sufficient to push both firms' investments in the "right" direction.[30]

[29] Supermodular complementarity, sometimes called Edgeworth complementarity, is the property that more of one asset or action makes more of the other more valuable. Market growth and cost-reducing investments are supermodular complements: a larger market makes the the lower costs more valuable, and lower costs make a larger market more valuable. Supermodular complementarity is sometimes offered as a reason for vertical integration, but, as the example shows, in a dynamic setting decentralized actors are capable of making coordinated investments. See Milgrom and Roberts (1990) and (1995) for more on the properties of supermodular functions.

[30] One might object by noting that, each firm receives only a fraction of the revenue from end-product sales thus will not push its investment as far as if the two firms combined. But in a dynamic context, each will get an extra kick from the investments of the other, thus over time, will make heuristic adjustments to their investment models.

The presence of bottlenecks scattered among the steps causes decentralized decision-making to break down and thus creates a "demand" to place all steps under common ownership.[31] For example, suppose Downstream sees an opportunity to increase the size of its market, but Upstream faces a bottleneck in production. Upstream will then not be able to increase its production to meet Downstream's new demand, and Downstream's investment will not pay off. Symmetrically, if Upstream invests in labor-saving equipment and Downstream faces bottlenecks in distribution, then Downstream will not increase the quantity of goods ordered. Upstream will gain from lower costs, but will not be able to grow, reducing its return on investment.

If one side attempts to contract with the other to fix its bottleneck, a classic holdup problem ensues. The bottleneck, by definition, constrains the throughput of the entire process. The owner of the bottleneck is thus in a position of great bargaining power vis a vis the owner(s) of the rest of the process. He or she has control of a unique asset on which the productivity of non-bottleneck assets depends. Under the standard reasoning of transaction costs economics, the owner of the bottleneck can expect to be paid a signficant percentage of the value gain to the entire process that comes from fixing the bottleneck. The expectation of holdup, in turn, reduces the value of systematic management to the owner(s) of the other steps.

In summary, separate firms carrying out interdependent step processes subject to bottlenecks have reduced incentives to invest in market-expanding or cost-saving technologies. *Even more insidious is the fact that neither will have incentives to invest in systematic management to identify and eliminate bottlenecks in its own processes.* If each is hostage to the other's bottlenecks, in equilibrium, neither will invest to increase its own throughput, for such investments will be wasted. In effect, non-integrated firms are in a classic coordination game with respect to investments in systematic management: both must

[31] A transaction free zone is a physical or virtual space where transfers of material, energy and information take place as simple transfers, without becoming transactions. Transactions are subject to mundane transaction costs of defining, measuring, and arranging for compensation, thus simple transfers are less costly than transactions. A corporation is by law a transaction free zone, although large corporations generally have accounting systems that provide for internal transactions between business units. See Baldwin (2008).

invest if either is to benefit.[32]

Placing Upstream and Downstream's processes within a single firm under common ownership changes their incentives to address bottlenecks. Within a vertically integrated firm, information can flow to a central authority who can take appropriate action without seeking permission of a third party. Bottlenecks can be addressed wherever they arise. A single vertically integrated firm under common ownership thus has greater incentives to invest in systematic management tools and techniques than separate firms carrying out the same step processes within a supply chain.

In fact, in the late 19th Century, across a range of industries, firms that vertically integrated and then rationalized their flow production systems came to dominate their non-integrated rivals through a combination of rapid market expansion and impressive cost reductions. The result was the emergence of a new class of organizations, which Alfred Chandler labeled "modern corporations."[33]

3.1 Hierarchy and Authority

The large modern corporations that emerged at the turn of the 20th Century were not only vertically integrated, they were also organized in a hierarchical fashion into functional departments. Each functional group was responsible for some portion of the flow process—procurement, fabrication of parts, assembly, distribution, and marketing. Proximate units were grouped together under the direct authority of a more senior official. These hierarchies generally extended from the very top of the company down to the level of front-line workers.

Two questions arise. First, why a hierarchy? Second, why direct authority? I will address each

32 Let V denote the value of systematic management if both firms invest, and let a denote the cost of such a program. Assume that $V >> a$, that is, systematic management is very profitable. The payoff matrix for the game is:

		Downstream	
		Invest	Don't invest
Upstream	Invest	V-a, V-a	-a, 0
	Don't invest	0, -a	0, 0

With simultaneous moves, there are two equilibria: both may invest or both may not invest. However, if at the starting point of the game, neither has invested, then the game is at an equilibrium and neither will invest.

33 Chander (1977).

question in turn.

Viewed strictly as a means of structuring communication linkages, hierarchy is an efficient way to filter large amounts of day-to-day, month-to-month, and quarter-to-quarter operational information. This was in fact the essence of James Thompson's theory of design for "boundedly rational" organizations: create groups that mirror the individual's need to communicate and coordinate actions in real time; then create groups of groups according to declining interdependency. The groupings fostered timely mutual adjustments and the hierarchy served as a means of conflict resolution "with each grade in the hierarchy specializing in resolving conflicts of the grade beneath it."[34]

In a flow process, the highest levels of interdependency arise between nearby steps, thus most of the necessary communication takes place locally. In a hierarchy, only selected information gets passed up to higher levels and then possibly back down to distant groups. The whole can be coordinated by setting consistent objectives for throughput throughout the whole organization within a given time period.

Groups at the lowest level of the hierarchy can manage to the plan, dealing with small deviations as necessary. This is "coordination by mutual adjustment."[35] Large deviations can be flagged and passed up to successively higher levels according to their magnitude. Such flags can cause additional problem-solving resources to be allocated to the point of disruption and might also trigger revisions in the plans of other departments. If the functional units are buffered from one another, for example by intermediate inventories, then the effect of a disruption in one segment will be attenuated in the more distant parts of the enterprise. Thus a hierarchy is an effective way to match the scope of communication and the scale of responses to the magnitude of random disruptions, wherever and whenever they arise.

In principle, however, none of the actions and responses described in the previous paragraph necessarily entails the exercise of direct authority. Direct authority involves control by one person of another persons actions: A gives orders and B obeys them. Granted someone must design a hierarchy,

[34] Thompson (1967) p. 60.

[35] Ibid. p. 56.

define the relevant groups and levels, and design appropriate incentives for each group and level. But these are all forms of indirect authority, which Charles Perrow called "controlling the premises of decision-making."[36] To avoid conflicts and ensure that all parts of the flow process receive attention (remember every step is necessary to successful completion), it may also be necessary to assign people to groups and levels. This is another form of indirect authority, which Perrow called "bureaucratic control."[37]

However a notable feature of modern industrial corporations in the U.S. was the fact that managers, sometimes through intermediaries such as foremen, exercised direct, "fully obtrusive" control over the flow of work and the actions of workers. Bosses gave orders and in general they were obeyed. Under the law, failure to obey an a boss's order was grounds for discipline or firing.[38]

3.2 What is Direct Authority Good For?

Charles Perrow has argued that there are three types of control in organizations: (1) direct, fully obtrusive control where orders are given and obeyed and performance is closely monitored; (2) bureaucratic control where people are assigned to specialized roles (tasks), which they perform under looser supervision; and (3) control of the premises of decision making where "the subordinate *voluntarily* restricts the range of stimuli that will be attended to … and the range of alternatives that would be considered."[39] The question is, what characteristics of a technological process make direct control necessary or useful?

I suggest that direct authority is good for three things: (1) synchronization; (2) education; and (3) coercion. Synchronization comes first. Coordinating actions precisely in time is a human skill, not found

[36] Perrow (1986) p. 129.

[37] Ibid. Some enterprises, notably open source communities, function well without bureaucratic controls. However software systems are generally made up of groups of that exist in stable form and do not need constant monitoring. Many of these modules are also non-essential, in the sense that the system as a whole can function without them. Step processes differ in that, by definition, all steps are essential and their completion requires explicit tasks to be performed in real time.

[38] Masten (2008); Freeland (2016).

[39] Perrow (1986) p. 129. Emphasis in original.

in other animals. But, especially when large numbers are involved, centralized authority is needed both to keep time (think of an orchestra conductor or the coxswain of a crew) and to design and assign the sequence of coordinated actions (think of a choreographer or an engineer laying out an assembly line). The more arbitrary the sequence, the more important it is to place a single decision-maker in charge.

In high-speed processes, any slowdown for consultation or negotiation will undercut the efficiency of the process and may destroy its effectiveness. Furthermore, if the process involves many people, it is useful to have someone outside the synchronized space, who can first determine a feasible and consistent set of actions (the "plan") and then identify points of imbalance (bottlenecks) and redress them. Thus it is best if direct authority is given to a designated actor (or actors) in advance. (Authority over different aspects of the process can be split up among different actors. Thus a conductor has authority over the tempo of the performance, while a choreographer has authority over the sequence of actions to be performed. Importantly, these actors exercise their authority at different times.)

Direct authority is also useful when one person knows more than another about the task at hand. A teacher can teach basic tasks and skills by giving orders, observing performance, and providing feedback and correction. When the tasks are programmed and the skills are physical, direct authority is an effective and natural method of instruction. Giving orders can be helpful and reassuring when a student truly does not know what to do.[40] It is only when trying to teach higher-level skills—e.g., judgment under uncertainty, problem selection, or emotional control—that the instructor needs to forgo direct authority for more indirect and unobtrusive methods of control.

Finally, direct authority can be a means of coercion, of making someone do something he or she would not choose to do on their own. A direct order backed up with enforcement can send someone into danger. It can put someone to work on boring, repetitive tasks. It can direct someone to perform actions that are meaningless in themselves and have no intrinsic value to the actor, or even are distasteful and

40 This is not to deny the fact that some humans are averse to any form of direct authority, hence resistant direct instruction. Herbert Simon has argued that such individuals are at a fitness disadvantage relative to those who are more receptive to social influences. See Simon (1990) and Augier and Simon (2003).

repugnant. Of course, bureaucratic controls and control of context can also be used coercively, but direct orders and surveillance may be the most efficient way to obtain unwilling or grudging compliance in the short run.

Given this profile, it is not surprising that some of the earliest and most successful uses of direct authority were in military settings. The Roman legions proved in their time that an infantry subject to direct authority and trained to carry out synchronized actions was militarily superior to loosely coordinated cavalry or disorganized hordes inspired by alcohol and plunder.

In the 18th and 19th Centuries, the principles of an authority hierarchy were carried over to commercial enterprises, beginning with cloth and lace factories, extending to railroads, and then to corporations engaged in mass production, such as steel mills, meatpackers and automakers.[41] All of these enterprises depended on synchronization of workflow for safety and efficiency. Their production processes were also technologically advanced, and thus beyond the comprehension of most workers. At the same time, the processes required human laborers to carry out precise tasks in a strictly timed order. The tasks themselves were often strenuous, boring, repetitive, and even dangerous. Incentives to shirk were high, but a combination of direct surveillance and output measurement could be used to maintain effort and throughput at acceptable levels.

The step-based production technologies of the First and Second Industrial Revolutions thus met all three criteria for the efficacy of direct authority. In their core production, distribution and marketing activities, most of the enterprises using the new high-flow-through technologies were organized as a hierarchy of "bosses" exercising direct authority over subordinates. Direct authority, bosses, and modern technology thus came to be seen as inextricably intertwined.

The technological requirements that made this organizational form efficient then faded into the background. It came to be taken for granted that authority was essential to the way modern firms worked.

41 Chandler (1977); Hounshell (1984); Landes (1986); Fields (2004).

However, in the late 20th Century, these assumptions were challenged in two ways: first, by the Toyota Production System which showed how flow processes might be managed more productively by engaging workers in a process of continuous improvement; and second, by the increasing importance of non-flow technologies especially in the realm of information goods and software.

4 The Rise of Corporations

Flow rationalization is a technical process that rewards certain organizational choices, specifically the inclusion of all potential bottlenecks in a single firm, a hierarchical organization structure, and the exercise of direct authority in the design of work and the supervision of workers. Organizations displaying all three properties became both common and powerful at the turn of the 20th Century. They became the exemplars—the poster children—of "modern" technology and "modern" organizational design.

The law regarding corporations, which took its current form at the end of the 19th Century, provided a legal framework that supported all of the properties demanded by flow technologies. By the 1850s, in most states, businesses organized as corporations had the ability to create zones of property ownership that reflected the interdependencies in the underlying technical processes and could last for indefinite periods of time. In contrast to proprietorships and partnerships, the technical processes carried out within a corporation would not be interrupted by the death or bankruptcy of an owner—they would continue under ownership of the corporation, and only the shares would change hands.[42] In addition, assets not essential to the the technical process could be placed outside the corporation's ownership: if the corporation subsequently failed, those assets could not be seized by the corporation's creditors (limited liability). Finally, beginning in 1889, when New Jersey passed a law permitting holding companies, a

[42] Hansmann, Kraakman and Squire (2006) call this feature "asset partitioning."

corporation's zone of activity could extend across state boundaries.[43] These were all new and valuable features of corporations relative to the preceding legal forms of business organization.

Legally constituted corporations are intrinsically hierarchical, since all decision rights are ultimately traceable to a single legal "person" whose actions are controlled by a Board of Directors.[44] This kernel of hierarchy could be elaborated into a hierarchical organization through the Board's power of delegation. However, delegated decision rights could be withdrawn at any time—they were "loaned and not owned."[45] Undelegated, residual decision rights were vested in the Board of Directors and could not be transferred without transferring the ownership of shares.[46]

The laws governing the relationship of employers and employees also evolved in ways that confirmed the corporation's direct authority and close control of work processes. Indeed employment law in the U.S. and Great Britain was essentially adapted from prior law governing master-servant relationships.[47] In contrast, contract law did not give managers rights of close control over the way contractors performed their work.[48]

It is no accident that modern corporations have exactly the properties needed to rationalize a multi-step process. There were enormous opportunities to create and capture value by rationalizing the

[43] Before 1889, states did not allow corporations operate outside their boundaries. Responding to this limitation, in the 1880s, companies in a number of industries combined to form "trusts." When trusts were formed, the entering partnerships had to incorporate so that their owners would have securities to exchange for trust certificates, thus the trust movement contributed to the rise of corporations. In 1889, New Jersey passed a law that permitted holding companies to own businesses in several states, and in 1890, Congress passed the Sherman Antitrust Act, which made the legality of trust agreements questionable. Thereafter, virtually all large enterprises were legally organized as holding companies. By 1910, corporations were the dominant form of organization for large businesses in the United States. (Navin and Sears, 1955).

[44] Blair (2003); Freeland (2016).

[45] Baker, Gibbons and Murphy (1999).

[46] Freeland and Zuckerman (2014).

[47] Coase (1937); Atleson (1983); Ahlering and Deakins (2007).

[48] There is a debate among law and organization scholars as to (1) whether the employment relationship differs from other contractual relationships in giving managers more control over how work is done; and (2) whether the employment relationship is an essential characteristic of firms. It seems fairly clear that, under U.S. law, an employment relationship provides managers with more direct control over work than a contracting relationship. However, it seems unnecessarily restrictive to say that employment relationships are a defining characteristic of firms. For example, many new ventures have no legal employees until they reach a certain scale. On this topic, see Alchian and Demsetz (1972); Jensen and Meckling (1977); Masten (1988); Freeland and Zuckerman (2014); and Freeland (2016).

many mechanized flow technologies that were invented in the latter half of the 19th Century. The limitations of the pre-existing legal forms—mainly proprietorships and partnerships, but also trusts and single-state corporations—could themselves be viewed as another set of bottlenecks reducing the efficiency of these processes. A great deal of money and effort was spent on contract redesign, lobbying of state legislators, and litigation in attempts to address these institutional deficiencies. In the end, the modern corporation emerged as the "winner" in competition with the other legal forms of organization. It was the legal framework actively chosen by the owners of large modern enterprises as most suited to their goals. Most of the companies that managed step processes began as proprietorships or partnerships, but, by 1917, virtually all of these opted to become corporations.[49]

The legal form was not without its problems and critics, however. In particular, the empowerment of managerial bureaucracies and the increasing separation of ownership from control created opportunities for rent-seeking, risk-shifting, and empire-building on the part of managers. Although the legal form enabled managers to pursue efficiency in flow processes, large corporations were only as efficient as competition required them to be.[50]

Furthermore, in many corporations, managers abused their rights of close control in service of a theory of machine-like efficiency in production flows. Inside these companies, the right of close control over employees was used both to collect information and to redesign jobs. Systematic and scientific management pointed managers towards defining and assigning tasks in ways that caused deskilling, physical hardship, and devaluation of the workers' cognitive and decision-making abilities. Distrust, resistance and outright hostility on the part of workers was the common result.[51]

The characteristic response by managers was to resort to authority and to fight any attempt to

[49] Berle and Means (1932); Navin and Sears (1955); Navin (1970); Chandler (1977) Appendix A; Rosenberg and Birdzell (1986) p. 220; Roy (1999).

[50] See, for example Berle and Means (1932); Jensen and Meckling (1976); Roe (1991;1996); Roy (1999); and the large literature on agency cost and corporate governance.

[51] D. Nelson (1974); Noble (1984); Halberstam (1986); Drucker (1993).

organize the labor force. Over time, inside contractors were replaced by employees. Labor organizers were summarily fired. A wide gulf opened up between a powerful and growing cadre of managers and an increasingly disaffected and distrustful workforce. Only later, with the advent of the Toyota Production System and other Japanese organizational innovations, did it become evident that even higher levels of efficiency could be achieved by making workers part of a system aimed at continuous improvement. However, that demonstration lay many decades in the future.

5 Platform Systems vs. Step Processes

Digital technologies, which lie at the core of modern, computer-based systems, differ from flow technologies in two important ways. First, a computer is both a *composite system* made up of six separable functional components plus software and a *platform* for performing calculations that are chosen by the user. The value of a platform lies in the range of things it allows the user to do. Furthermore, in composite systems, the *options* to add new functionality and to substitute superior components for inferior ones can substantially increase the value of the system as a whole. Users interact, not with goods that perform pre-programmed tasks, but with evolving systems they partially design themselves.

Second, most of the hardware in computer systems came to depend on integrated circuits manufactured via the planar process. Integrated circuits had remarkable scaling properties, which allowed them to become ever smaller, cheaper and faster in each successive generation. However, because the computers were composite systems, different components changed at different rates. Each component depended on different ancillary technologies, such as electro-mechanical assemblies, vacuum tubes, X-rays, magnetism, as well as the new sciences of coding, algorithms, and programs. These disparate technologies progressed along separate trajectories and thus the different parts of a computer were capable of changing independently and asynchronously.

5.1 The Value Structure of a Platform System vs. a Step Process

The value structure of a platform and complements is very different from that of the step processes described in previous sections. In a step process, all steps are essential, and value is constrained by the step with the minimum capacity—the production bottleneck. The value of the process is proportional to its throughput:

$$V(\text{Step Process}) \text{ is proportional to } Q_{\min}(N) \equiv \min(q_1,...,q_i,q_j,...,q_N) \quad (4)$$

where $Q_{\min}$ denotes the expected capacity of the system defined as the minimum throughput of N steps, each of which is essential to the finished good. The goal of systematic management in a step process is to increase flow through the production bottleneck in order to increase the throughput of the entire process.

In contrast, a platform system consists of a core set of essential components plus a set of complements. The user of the system must have the platform in order to take advantage of the complements. However, unlike the steps in a flow process, each complement is *optional*: if it is absent, the platform and other complements can still function. Thus for each complement, the user of the system can assess whether its value exceeds its cost. If the complement passes this test, the user will add it to the system, if not, it can be left out. The value of a platform system is proportional to its options:[52]

$$V(\text{Platform System}) \text{ is proportional to } P \cdot [E\max(a_1,0;k_1) + ... + E\max(a_j,0;k_j) + ...] \quad . (5)$$

Here P is a binary variable indicating the presence or absence of the platform. Each term within the square brackets denotes the expected value of the *maximum* of k independent draws of a random variable a and zero. In other words, the j^{th} option is the site of k_j independent experiments and the user selects the experimental outcome with the highest value.[53] In the presence of the platform ($P=1$) the value of the system is the *sum* of the values of the individual options.

Optional complements include components that add new functionality to the system, for example,

[52] Baldwin and Clark (2000) p. 264.

[53] Users may place different values on experimental outcomes, in which case they may make different selections.

new games in a video game system, or new software applications and hardware in a computer system. Optional complements *also* include optional features and upgrades of parts of the platform.

Three propositions based on this value structure show how the technological requirements of platform systems differ from those of step processes.

First, the presence of an option can never decrease the value of a system. This is apparent from equation (5). If the best version of a particular optional complement degrades the system or is simply not worth the cost, then $\max(a,0;k)$ will be zero. The user will simply not incorporate that option into his or her system. From this we obtain:

Proposition P-1 (Positive Impact of Options). The more options associated with a platform system, the greater the value of the system.[54]

Second, in striking contrast to step processes, increasing the risk (variance) of any option outcome does not harm the system, and may increase its value. This is a well-known property of options. Intuitively, the outcome of any random draw from a probability distribution can be rejected if it is less than zero (or, in the case of multiple draws, if it is less than the maximum of all other draws). Increasing the variability of the gamble increases value because the risk-taker is shielded from bad outcomes. This leads to:

Proposition P-2 (Positive Impact of Risk). In a platform system, the greater the variability in the value of any optional complement, feature, or upgrade, the greater the value of the system.[55]

---

[54] Proof is immediate.

55 **Proof of Proposition P-2.** The proof is essentially parallel to that of Proposition S-2 above, except that focal function is convex, not concave. Again, consistent with Rothschild and Stiglitz (1970), I define increasing variability (risk) as the addition of a mean preserving spread to a given probability distribution. Consider one of the optional components whose expected value value is the maximum of *K* variants and zero:

$$E\max(a,0;k)$$

Max(*a*, *0; k*) is a convex function, thus, as demonstrated by Rothschild and Stiglitz:

$$E\max(a+\varepsilon,0;k) \geq E\max(a,0;k)$$

where ε is a mean preserving spread. Therefore increasing the variability in outcomes for any option weakly increases the expected value of the option and thus the entire system. *QED.*

Finally, a key result obtained in *Design Rules: Volume 1* was to show that dividing an option into independent gambles (while preserving the variance of the sum) increases the value of the system. This proposition is derived from the insight, originating with Robert Merton, that a "portfolio of options" is worth more than an "option on a portfolio."[56] This in turn implies:

Proposition P-3 (Power of Modularity). In a platform system, dividing any component into modules that can be developed independently and mixed and matched after the fact increases the value of the system.[57]

Proposition P-3 is in fact a corollary of Proposition P-1: by subdividing the system into modules, the architect increases the number of options, increasing the value of the system. To make the argument more concrete and intuitive, consider two computer systems, each made up of four components: a drive system, a main board, an LCD screen and packaging. Design work takes place to improve/upgrade each component. For simplicity, assume that, for each component in each system, there is a 50-50 chance that the new design will be better (=+$20) or worse (=–$20) than the previous design.

System A is designed as an integral system, in which the component designs are interdependent and cannot be split apart. System B is designed as a modular system with design rules that split up the components and allow prior designs to be retained if the new designs are inferior.

Figure 5-1 shows one possible outcome for the two systems. Here, in each case, the drive system and LCD screen designs turn out to be better (+$20) than previous designs while the main board and packaging designs are worse. The integral system imposes an "all or none" constraint on the options. Because the component outcomes are mixed, the integral system is worth no more than the system it is meant to replace (*Value = 0*). In contrast, the modular system allows designers (or users) to reject the inferior component designs, selecting only the superior solutions. The modular system is thus more

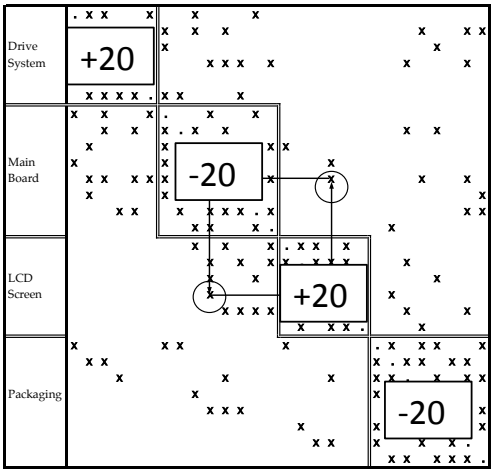
[56] Merton (1973).

[57] Baldwin and Clark (2000), p. 259-264.

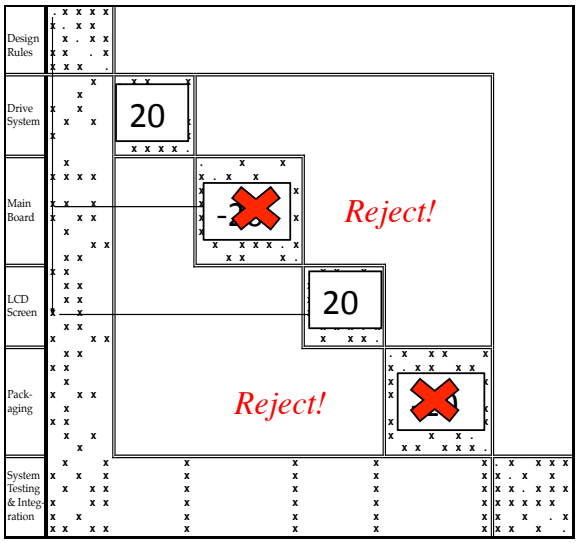
valuable than the integral system (*Value = 40*).[58]

Figure 5-1 Contrasting Values for an Integral and Modular Systems Given Risky Component Outcomes

(A) Integral Design
System cannot be split apart
Take all or nothing
Value =0



(B) Modular Design—*Modules are separate options*
Value = 40



5.2 Platform Systems and Step Processes Compared

We are now in a position to compare the technological dimensions of platform systems with step processes. The dimensions I consider are (1) the impact of risk on value; (2) the pattern of technical dependencies; (3) the value of modularity. The contrasting properties of of platform systems and step processes in turn suggest that different organizational forms are needed to exploit these different technologies.

Impact of Risk. An option protects the user and the system from downside risk, thus increasing the variance of outcomes for an option increases its value. Platform systems whose value is derived from options thus support risk-taking with respect to complements, features and upgrades. Platform managers

[58] This reasoning can be generalized to any set of underlying probability distributions as long as there is some degree of independence in the component outcomes for both the integral and modular systems.

can encourage many crazy departures from the status quo, as long as they protect the platform itself from harm. In contrast, in a step process, no step is optional, and the worst-performing step (the production bottleneck) constrains the process as a whole. Thus variability in step outcomes and risk-taking within steps decreases value.

Patterns of Dependency. Platform systems and step processes also have very different patterns of technical dependency. Ideally, platforms are modular systems, with complements, upgrades and features depending on platform design rules but not on each other. Eliminating dependencies across components creates additional options for users, thus increases the value of the platform system. In contrast, all steps in a step process are interdependent. Flow efficiency is maximized when all steps have the same throughput in each time period. A disruption or change in throughput in any step requires adjustment of all other steps. And as one production bottleneck is resolved, another appears somewhere else in the process. (Steps can be insulated by creating buffers: in effect, buffers modularize the step process.)

Value of Modularity. Modularity in a platform system can be created by first understanding the physical interactions between elements *in detail*. Lateral dependencies requiring real-time coordination can then be eliminated and replaced with hierarchical design rules.[59] In this fashion, the platform can be separated from the optional complements. The platform itself and each complement can be further modularized to support the addition of optional features and upgrades.[60] The risk in this process is that, when lateral dependencies are not well understood, premature modularization runs the risk of system failure.[61] Thus, especially in new systems, the option value of mix-and-match flexibility must be weighed against the possibility that unknown dependencies between supposed modules may appear, leading to coordination delays and even system failure.

Modularity in a step process can be created by placing buffers—either stocks of partially

[59] See Baldwin and Clark (2000), Chapter 3, for details.

[60] Garud and Kumaraswamy (1995).

[61] Colfer and Baldwin (2016) p. 720.

completed goods or time lags—between steps. Buffering insulates downstream steps from upstream variation. Steps that become temporary production bottlenecks have time to catch up and replenish stocks without disrupting downstream operations. As a result the process as a whole becomes more resilient. However, buffering increases the need for capital (inventory) and the time needed to complete the process. Thus the value of resilience achieved through modularity must be weighed against the loss of throughput and efficiency.

Table 5-1 summarizes the main technological differences between platform systems and step processes as derived from their contrasting value structures.

Table 5-1 Contrast between Platform Systems and Step Processes

Dimension	Platform Systems	Step Processes
Risk	Variability within options *increases* value.	Variability within steps *decreases* value.
Dependence	•Complements, upgrades and features depend hierarchically on the platform. Absent the platform, they have no value. •Complements, upgrades and features that are modules can be chosen independently. Choosing one does not require or prohibit choosing another. •Variants of modules may be mutually exclusive, e.g., a car can be black or red, not both.	•Steps are interdependent. All steps are needed to complete the product. •Bottlenecks are interdependent. Fixing one creates another.
Modularity	•Modularity in a platform system can be created via design rules governing the architecture, interfaces and tests. •Modularity creates options, increasing the value of the system. •The option value of modularity must be weighed against system failure caused by unknow, cross-module dependencies (premature modularization).	•Modularity in a step process can be created via buffers. •Modularity insulates steps from upstream disruptions making the process more robust, but less efficient. •The value of modularity must be weighed against a loss of throughput efficiency.

5.3 Organizational Implications

We come now to the organizational implications of the technological differences between platform systems and step processes. Above, I argued that the rationalization of a step-based production process using the tools of systematic management is most efficiently conducted within a *single, vertically integrated firm* that spans all potential production bottlenecks. In addition, in a large enterprise, a nested *hierarchy* of groups is an efficient way to organize information flows. Managers at the top of the hierarchy can set consistent plans for all organizational units. Managers at lower levels can filter information and address local deviations from plan. Finally, in the short run, *direct authority*—the ability to give orders and have them obeyed—is an efficient way of implementing the changes in job content and work flow needed to address bottlenecks.

In contrast to step processes, platform systems benefit by multiplying options, increasing the variability of option outcomes, and modularity. By definition, the modules in a platform system do not need to synchronize tasks or coordinate decisions: interoperability is ensured by adherence to design rules. In this respect, platform systems are more "loosely coupled" than interdependent step processes. As a result, the modules of a system need not be grouped within a single firm. The boundaries of modules offer thin crossing points in the task network, where transaction costs are low.[62] Thus once the platform and options have been split into separate modules, different components can be supplied by different firms with no loss of interoperability or efficiency.

In effect, the contrasting technical architectures of platform systems and step processes drive organizations in opposite directions. The over-riding mandate in a step process is that all steps be performed in a predictable fashion with no bottlenecks to constrain output. Uncertainty is the enemy and the job of managers is to eliminate variation. In contrast, the mandate in a platform system is to provide users with options they can exercise at will. The more uncertain users are about what they will want, the

[62] Baldwin (2008).

more valuable are the platform's options. Platforms are particularly valuable when users cannot envision how their future problems will be solved but believe that new solutions will appear via the platform in the normal course of events.

Because a platform and related options can be supplied by many different organizations, it generally does not make sense to speak of a single platform "owner." Below I will speak of the "platform architect(s)" and the "platform sponsor(s)." Platform architect(s) specify the platform's design rules—the architecture, interfaces, and tests that ensure the interoperability of platform components. Platform sponsor(s) exercise control over the design rules. Platform sponsors are often organizations, including for-profit firms, standards-setting bodies, and open source communities. In general, platform architects work for or on behalf of platform sponsors.[63]

5.4 The Tradeoff between Option Value and Flow Efficiency

Up to this point, I have treated platform systems and step processes as mutually exclusive technologies that offer different incentives and provide different rewards to organizations. Step processes respond to systematic management aimed at eliminating bottlenecks. Platform systems reward risk-taking and the generation of new options.

However, at a deeper level, platform systems and step processes are intertwined. Platforms and options do not arise out of thin air. Between an imagined technological possibility and a "real" technological choice, lies a technical recipe. A sequence of steps, whether long or short, must be carried out. *In other words, platforms and options are brought into the real world via step processes.*

Module boundaries determine what steps will be performed within a given module. Steps within modules are, by definition, highly interdependent; steps in different modules are (nearly) unconnected except for their adherence to a common set of design rules. More precisely, step processes within modules

[63] My use of the term platform "sponsor" is consistent the definition of Van Alstyne, Parker, and Choudhury (2016). Gawer and Cusumano (2002) use the term platform "leader" to refer to this role. Moore (1996) defines an "ecosystem leader" as a firm that enables the members of a shared ecosystem "to move toward shared visions to align their investments" (p. 26). Iansiti and Levien (2004) define a "keystone" firm as one that controls "key hubs" in a business ecosystem and manages its position to promote the long-term profits or "health" of the network. Of these terms, "sponsor" seems the most neutral.

are separated from one another by thin crossing points in the task network. If transfers between two subsets of steps (i.e., tasks) are dense and complex, the tasks can no longer be considered to be in separate modules since changes in one set will necessitate changes in the other.[64] Therefore the boundaries of modules and the scope of the underlying step processes are two faces of the same coin. And in the process of designing the breakpoints between modules, option value must be weighed against flow efficiency.

Where should one set boundaries between modules, which are also necessarily the boundaries of the underlying step processes?

The first consideration is the state of the designers' knowledge about the underlying technology. If the technology is well-understood, then architects of the system can replace real-time problem-solving focused on resolving technical dependencies with rules that ensure compatibility between discrete components. Ignorance about physical and logical interactions between components in a technical system is thus a paramount reason to place tasks related to those components within the same module.

Efficiency and lower cost are also reasons to tie tasks together in a synchronized step process within a module. For example, in 19th Century steel-making plants, there were large savings to be gained by transferring molten pig iron from a blast furnace to a Bessemer converter and further savings in transferring molten steel from the converter to a rolling mill. In the earliest steel mills these three stages were separate modules. Later experiments aimed at increasing throughput led to the invention of new machines such as the Jones mixer and the Wellman charging machine, which allowed steel makers to tie the steps together to achieve a continuous flow of metal. Similarly, the components of automobiles were initially made in different shops. When Henry Ford and his managers invented the moving assembly line, they realized significant savings by tying different stages together in a continouous flow process.

Thus the efficiencies achievable through synchronized flow are another reason to place tasks in the same module. However, the choice of breakpoints in the process depends on the relationship between

[64] Baldwin (2008).

costs of production vs. the option value of of changing parts of the process after the fact. When options to change the process piecemeal are valuable *then it makes sense to sacrifice some amount of flow efficiency to "expose" the options and make them more easily available*.

In fact, this is the lesson General Motors taught Ford in the 1920s. Ford optimized its production system for flow efficiency and thereby achieved very low costs per vehicle. But it offered customers very few options. Among the things Ford did *not* incorporate in its cars were innovations that improved the ease of driving, comfort and style—things like automatic transmission, electric starters, shock absorbers, cushioned seats, and colors. GM, in contrast, offered customers a range of cars and introduced new features and styling in every model every year. GM's production lines were less efficient and GM's cars cost more because of the variety and features offered. However, in the end, the value of the options offered to users more than made up for any increase in production costs.[65]

Moore's Law—the prediction that chip densities would double and costs fall by half every eighteen months to two years[66]—similarly affected the trade-off between flow efficiency and modularity in the industries that used semiconductor chips. With each new generation, the number of possible chip designs expanded. The number of users who could afford sophisticated computers also went up as prices went down. As the number of users increased, the number of things they wanted to do with their computers, mobile phones, notebooks, tablets and other devices increased as well.

The dynamics of Moore's Law thus gave rise to an exploding set of potential options to add new components and to upgrade older components in computer and communication systems. Rewards to modularizing both hardware and software increased in line with the demand for new functions, features and upgrades. As a result, modular step processes ceased to be oxymoronic and became common. Step processes that could be quickly set up, modified, and dismantled were preferable to efficient but inflexible

[65] Abernathy, Clark and Kantrow (1983); Clark (1985); Hounshell (1985).

[66] G.E. Moore (1965; 1975); Mollick (2006).

processes that delivered standardized products in large volume.[67]

The exceptions to this general trend were microprocessors and DRAMs. These standardized chips were manufactured in large volumes using intricate step-based production processes. Much like Lego bricks, they became the fundamental building blocks for modular computer and communication systems. However, the chips themselves were highly integrated with many interdependencies among their design elements. And the corresponding production processes, especially fabrication, were also highly integrated with many interdependencies between steps.[68]

5.5 Capturing Value in a Modular System: The Problem of Exclusion

From the perspective of a platform architect or sponsor, optional complements should be separated from the platform and from each other. In many cases, components within the platform itself can also be divided into separate modules and supplied by third parties. This means that a platform architect or sponsor cannot separate the pursuit of platform *options*, which are the platform's main source of value, from the question of platform *openness*, that is, who can attach their products to the platform and on what terms? Modularity multiplies options and can create large amounts of value for users. But it also provides a means for third parties to enter the system by supplying modules.

IBM, the sponsor of System/360, discovered this fact in the early 1970s when makers of plug-compatible peripherals introduced storage and memory devices (and eventually processors) that were fully interchangeable with similar units sold by IBM. Users could incorporate these non-IBM devices into their IBM systems. The plug-compatible devices were always cheaper and often faster than comparable IBM products.

A modular architecture necessarily creates thin crossing points in the task structure of the underlying technical system. Thin crossing points have low transaction costs. The creator of the modular system can use the thin crossing points as points of attachment to sell optional products and upgrades to

[67] Sturgeon (2002); Berger (2005).

[68] Hilton (1998); Chafkin and King (2016).

its customers.

But, at the same time, third parties can use the same thin crossing points as points of entry for their products. They do not have to design and build an entire system; they only need to design and build a particular module. Thus the architect of a modular system will generally find itself facing competition from external suppliers of modules. Such competition benefits the purchasers of the modular system, thus customers can be expected to encourage the new entrants.

The likelihood of unauthorized entry through modules is affected by the status of the design rules established for the modular architecture. If the design rules can be protected by secrecy and/or intellectual property rights then the architect may be able to restrict entry to its own employees, suppliers, and licensees. However, the design rules for System/360 were known to many employees and most were simply protocols that did not involve a novel or non-obvious solution to a technical problem.[69] Although they held the system together, the rules themselved could not be patented.

IBM sued some plug-compatible peripheral companies for theft of trade secrets and was countersued for violations of antitrust law and predatory pricing. In the end, IBM's intellectual property claims were mostly upheld, but the damages assessed were small, and many peripheral manufacturers went on to become established firms.[70]

Nevertheless, IBM's managers found that if they controlled the evolution of a modular system, they could tolerate and even encourage entry in some parts of the system, and the entire line of products would still be very profitable.[71] The presence of complements could increase total demand, and IBM might sell more processors and peripherals as a result. Most IBM managers did not think this was a good way to build a business, but in the late 1970s, at least two—William Lowe and Don Estridge—began to see the possibilities. They in turn had a direct line to the chairman of IBM, Frank Cary, who wanted the

[69] Pugh, Johnson and Palmer (1991) Chapter 9; DeLamarter (1986).

[70] Baldwin and Clark (2000) Chapters 14 and 15.

[71] Ferguson and Morris (1993).

company to enter the new, high-growth market for small computers. In this way the stage was set for the introduction of the IBM PC, *the first radically open, modular computer system*.

6 Open Platform Systems

The last two decades of the 20th Century saw the rise of three distinct types of open platforms and surrounding ecosystems. First, reflecting the fact that all computers are fundamentally platforms for performing calculations and displaying results, numerous companies in the computer industry adopted a "forward open" stance and invited downstream complementors to create optional modules for their systems. Second, firms making physical goods broke up their vertically integrated production systems and created "backward open" modular supply networks. Finally, in the 1990s, the Internet and the WorldWide Web led to the creation of numerous "open exchange" platforms—websites designed specifically to facilitate transactions and other valued exchanges of goods, information, and opinion.

The three types of open platforms systems were similar in many ways. All were based on a fundamental modularization between the core platform and optional components. All benefited from modularity in the optional components and often in the platform itself. All relied on design rules—an architecture, interfaces and tests—to ensure interoperability. Finally, all three types of platform supported the decentralization of tasks and decision-making power across different firms, organizations and individuals.

None of these platform types was truly new. Precursors of forward-open platforms include infrastructure such as the electrical grid, water distribution systems, the railway network and road systems going back to ancient times.[72] Before the advent of large synchronized flow systems in manufacturing, the production and distribution of goods was organized as an open modular supply network.[73] Finally,

[72] Frischmann (2004; 2012) defines infrastructure as a capital resource that provides opportunities (options) to many actors, and whose value lies in "downstream productive activities." He explicitly identifies infrastructure with platforms: "Essentially, infrastructure resources are enabling "platforms" on which others build" (2004, p. 957).

[73] Chandler (1977) Chapters 1 and 2. Rosenberg and Birdzell (2008) Chapter 5.

telegraph and telephone networks, commodity and stock exchanges, and marketplaces existed as open exchange platforms well before the advent of the Internet.[74]

What was new were the things made possible by electronic and digital technologies. First, broadband communication took place at the speed of light, and thus the new digital platforms had global reach. Second, the fundamental physical entities (chips and circuits) were subject to ongoing miniaturization under the metronome of Moore's Law. Because the physical devices were small, they could be fabricated in one place, assembled in another, and shipped to yet another at low cost. Third, because designs were ever-changing, the value of modularity in both products and processes was high. (However, the trends of miniaturization and modularity worked in opposite directions. Miniaturization made it possible to pack more circuit elements on a small chip, but the chip itself was an indivisible module within the larger system. Thus Moore's Law turned single chips into ever larger, more complex modules.)

Last but not least, from an organizational standpoint, digital platforms could be made virtual. That is, through modularization, the control of *critical visible information* that established standards of interoperability could be separated from control of physical assets. This meant that entrepreneurial startups with few physical assets but a superior understanding of technology, could aspire to become platform sponsors.

6.1 The Evolution of Open Platform Systems in the Computer Industry

As a matter of historical record, the evolution of open platforms in the computer industry followed a clear trajectory. Although by definition all computers are platforms for calculations, open platforms require highly modular architectures. The interface between platform and options must be clear and sharp and the options themselves must be encapsulated and relatively small. Early computers were

[74] Rochet and Tirole (2003); Boudreau and Hagiu (2011).

not modular. The first modular computer system was IBM System/360, introduced in the mid-1960s.[75] System/360 was meant to be a closed platform, but its popularity and the transparency of its interfaces allowed "plug-compatible" manufacturers to attach their products to the system at user sites without IBM's permission. Thus System/360 became open despite IBM's strong resistance, expressed in lawsuits and in a number of "tricks" the company played with prices, contract terms and technical interfaces.[76]

Because most startups lacked the resources to be vertically integrated, new entrants to the computer industry in the 1970s, including DEC and other minicomputer makers, were generally forward-open and backward-open to some degree. The IBM PC revealed just how far openness could be taken, as well as the competitive advantages of this strategy. However, in short order, the reverse engineering of the PC BIOS, the entry of numerous PC-compatible clones, and IBM's subsequent loss of market share and profitability, demonstrated the pitfalls of openness.

Thereafter, from the 1980s through the mid-1990s, firms in the computer industry experimented with different combinations of forward and backward openness. Then, following the the rise of the Internet in the mid-1990s, platforms dedicated to exchanges of goods, information, and opinion took center stage. Most firms that had previously sponsored forward-open or backward-open platforms incorporated open exchange platforms into their organizational designs.

The three types of open platforms—forward, backward, exchange—can be combined in different ways. Each type rests on a separable set of activities, and poses a different set of strategic challenges and risks. Thus, if we accept Alfred Chandler's argument that the primary challenge for managers in the late 19th and early 20th Centuries was to set up administrative systems that could efficiently supervise multi-step production processes, then a new challenge for managers in the 21st Century is to coordinate multiple interacting platforms and their surrounding ecosystems.

In this fashion, new digital technologies are reshaping organizations. At the same time, step

[75] Ferguson and Morris (1993); Baldwin and Clark (2000).

[76] Baldwin and Clark (2000) pp. 388-390; DeLamarter (1986).

processes have not disappeared. However, as I've argued, if the benefits of innovation and flexibility are high, step processes may be carved up into modular subprocesses within platforms.

7 The Vertical-to-Horizontal Transition in the Computer Industry

A modular technical architecture is a necessary pre-condition for an open technical platform. Without a high degree of modularity, there are by definition very few thin crossing points in the task network at which to place low-cost transactions, few points of attachment for would-be complementors, and few opportunities for value-increasing exchanges between members of an ecosystem.

By definition, a closed platform gives rise to a vertically integrated firm: the underlying architecture may be modular, but components and complements are made inhouse. IBM's radically open PC platform was the largest and most visible of the open platforms, but many other firms in the 1980s and 1990s sponsored open platforms or joined them as suppliers or complementors. *The vertical-to-horizontal transition in the computer industry could not have occurred were it not for open platforms.*

An open platform requires an ecosystem of suppliers and complementors. Opening a system in the absence of any external providers is like having a party with no guests. However, as indicated, the modular architecture of IBM System/360 provided many opportunities for module makers, both large and small, to enter the industry. Also, between 1976 and 1980, a new group of firms emerged that made processors, peripherals and software for microcomputers. Microcomputers were the fastest growing sector in the computer industry, with growth rates approaching 50% per year. The makers of plug-compatible peripherals and microcomputer parts were the basis for the business ecosystem that formed around the IBM PC platform.

7.1 Horizontal Layer Maps

A business ecosystem will naturally become organized according to the specific functions that each product or service fulfills within the overall system. Treating functions as categories, each separate module in the architecture can be thought of as occupying a different "layer" in a vertical "stack" of

functions. Firms in the ecosystem can then be associated with one or more layers in accordance with the functions their products perform.[77] For example, Firm A might make chips, Firm B might make storage devices, and Firm C might develop software. Each firm, along with its competitors, would appear in a different layer of the computer "stack."

Grouping products and firms by function, it is possible to construct a "layer map" of a business ecosystem.[78] Firms making most functional components inhouse will appear in several layers, thus forming vertical columns. Firms specializing in a single functional component will appear in only one layer forming part of a horizontal band. If open platform architectures and their ecosystems become more important relative to vertically integrated firms, the vertical columns will shrink and the horizontal layers will expand.

As discussed in the introduction, according to Andy Grove, the computer industry went through a vertical to horizontal structural transformation between 1980 and 1995. This transition was one of the key "surprises" in the evolution of the industry, something almost no one predicted or anticipated. Grove called it a "strategic inflection point" for the industry "when the balance of forces shifts … from the old ways of doing business and the old ways of competing, to the new."[79] In his words;

> Even in retrospect, I can't put my finger on exactly where the inflection point took place in the computer industry. Was it in the early eighties when PCs started to emerge? Was it in the second half of the decade, when networks based on PC technology started to grow in number? … [What is clear is that] by the end of the 1980s, many large vertical computer companies were in the midst of layoffs and restructuring … . [At] the same time, the new order provided an opportunity for a number of new entries to shoot into preeminence.[80]

With Michael Jacobides and Reza Dizaji, I used segment data on firm market values to construct a series of layer maps of the greater computer industry.[81] Plates 1-4 present layer maps showing the

---

[77] Providing a function can be thought of as performing a role in an industry architecture that mirrors the technical architecture. Jacobides, Knudsen and Augier (2006).

[78] Fransman (undated).

79 Grove (1996) p. 33.

[80] *Ibid.* pp. 44-45.

[81] Jacobides, Baldwin and Dizaji (2007).

market capitalization of the largest 14 firms in the greater computer industry relative to the whole industry in four different years: 1985, 1990 1995 and 2000.

Grove's perception of a vertical-to-horizontal industry transition is borne out by the maps. In 1985, IBM, the quintessential vertical, accounted for more than half the market value in the industry. Other vertically integrated systems makers, including the Japanese firms, Hitachi and NEC, and U.S.-based Hewlett Packard and Digital Equipment Corporation occupied the next four places. As Grove observed, "Going into the eighties, the old computer companies were strong, growing and vital."

Between 1985 and 1990, IBM, although still largest, lost a great deal of its market value, as did Digital, National Cash Register, Sperry, Unisys, and Wang Labs. However, the Japanese verticals, now joined by Toshiba held their own. Microsoft, which went public in 1986, occupied the #5 position, with Intel at # 6. Compaq, Novell, and Sun Microsystems joined Apple as makers of small computers systems and software. Packaged software (a new layer) accounted for over 10% of the industry's market value. Nevertheless the industry as a whole was still dominated by vertically integrated firms.

Between 1990 and 1995, the map changed dramatically. First of all, the industry as a whole greatly expanded in terms of total market value. Microsoft was now #1; a shrunken IBM is # 2; and Intel is #3. By this point, IBM had lost control of the PC platform and Microsoft and Intel had taken its place as platform sponsors. All the horizontal layers, but especially software, expanded at the expense of the verticals. Horizontal layers now accounted for around three-quarters of the industry's value. New entrants to the top tier included Cisco, Oracle, First Data Corp, CA Inc. and Micron Technologies replacing NCR, Digital, Apple, TI, and Novell. Also notable is the increase in "white" space in the map, that is, publicly listed firms that were not in the top fourteen. Industry concentration diminished as hundreds of firms entered the industry and went public. The early 1990s marked the beginning of the Internet Gold Rush, which turned into the Internet Bubble and Crash. (The Internet, of course, is an open platform, that dwarfs all previous open platforms in terms of the scope and diversity of it complements.)

By 2000, the verticals had disappeared. Following a major restructuring, a diminished IBM (# 5)

no longer claimed to be vertically integrated, but was concentrated in systems, services and software. Reflecting the importance of the Internet as a new platform, Cisco has moved into the # 1 position. New members of the top tier included EMC, Lucent Technologies, Dell, Taiwan Semiconductor, Juniper Networks, and STMicroelectronics. The three Japanese verticals—Hitachi, Toshiba, and NEC— as well as Compaq, First Data, CA Inc., Sun Microsystems, and Micron Technologies were off the list. All others again accounted for about half the industry's market value.

Thus a vertical-to-horizontal transition in industry structure did take place, beginning in the late 1980s and continuing through the 1990s and early 2000s. It could not have occurred in the absence of highly modular technical architectures and open platform systems. Modular architectures made open platforms feasible; open platforms with their myriad of suppliers and complementors then proved to be highly competitive against closed platforms. Over a twenty year timespan, vertically integrated firms essentially disappeared.

Plate 1 Distribution of Computer Industry Market Capitalization by Layer 1985

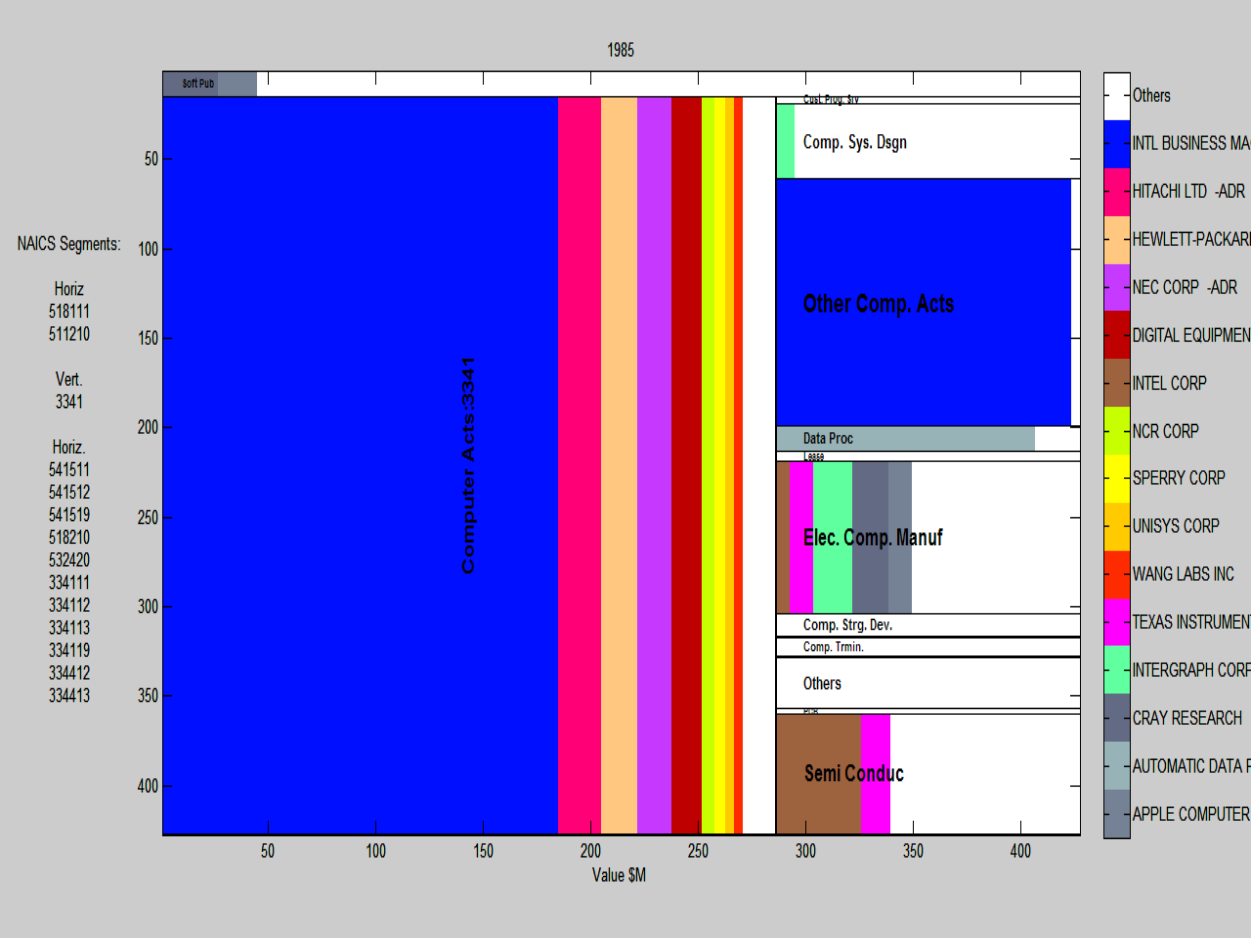


Plate 2 Distribution of Computer Industry Market Capitalization by Layer 1990

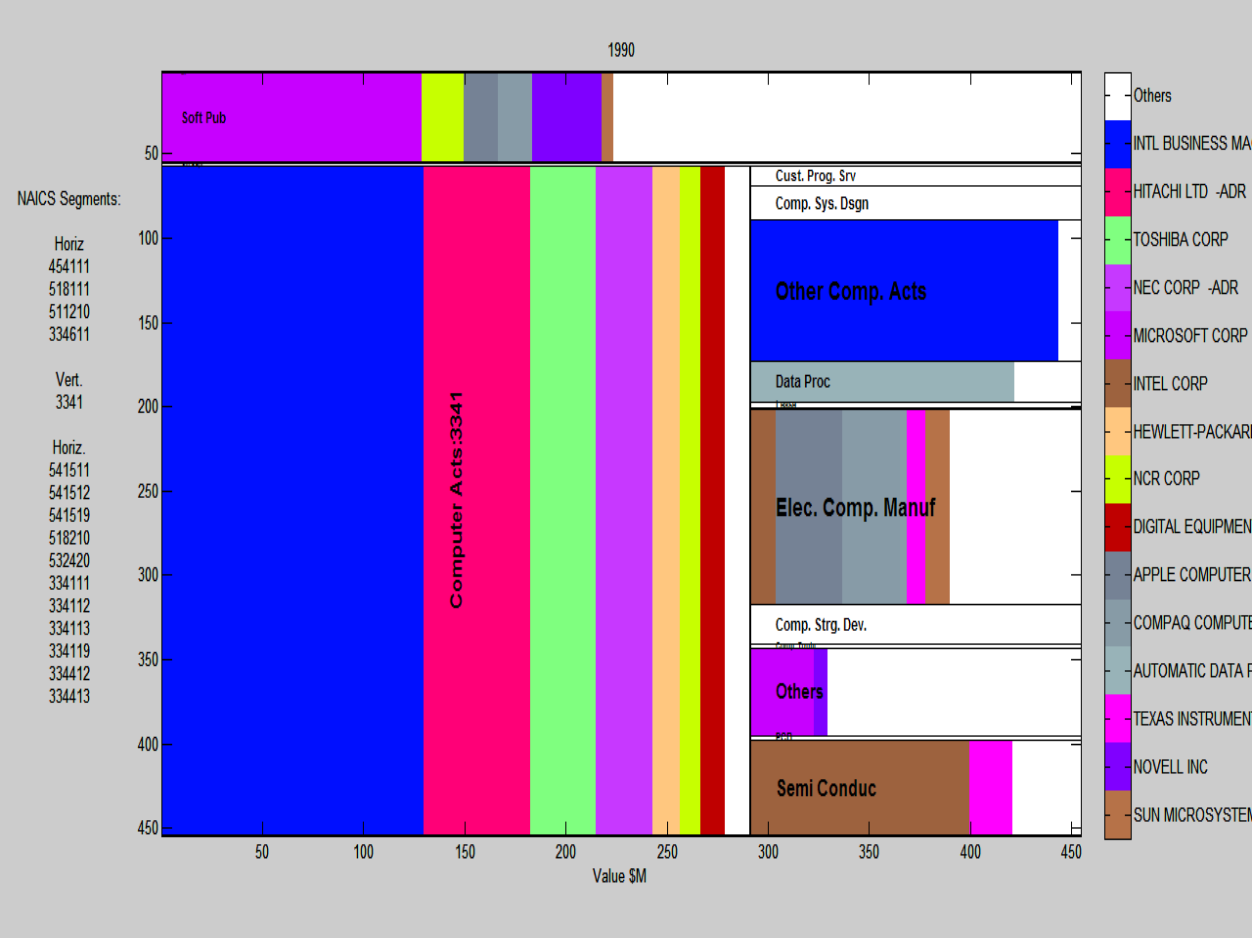


Plate 3 Distribution of Computer Industry Market Capitalization by Layer 1995

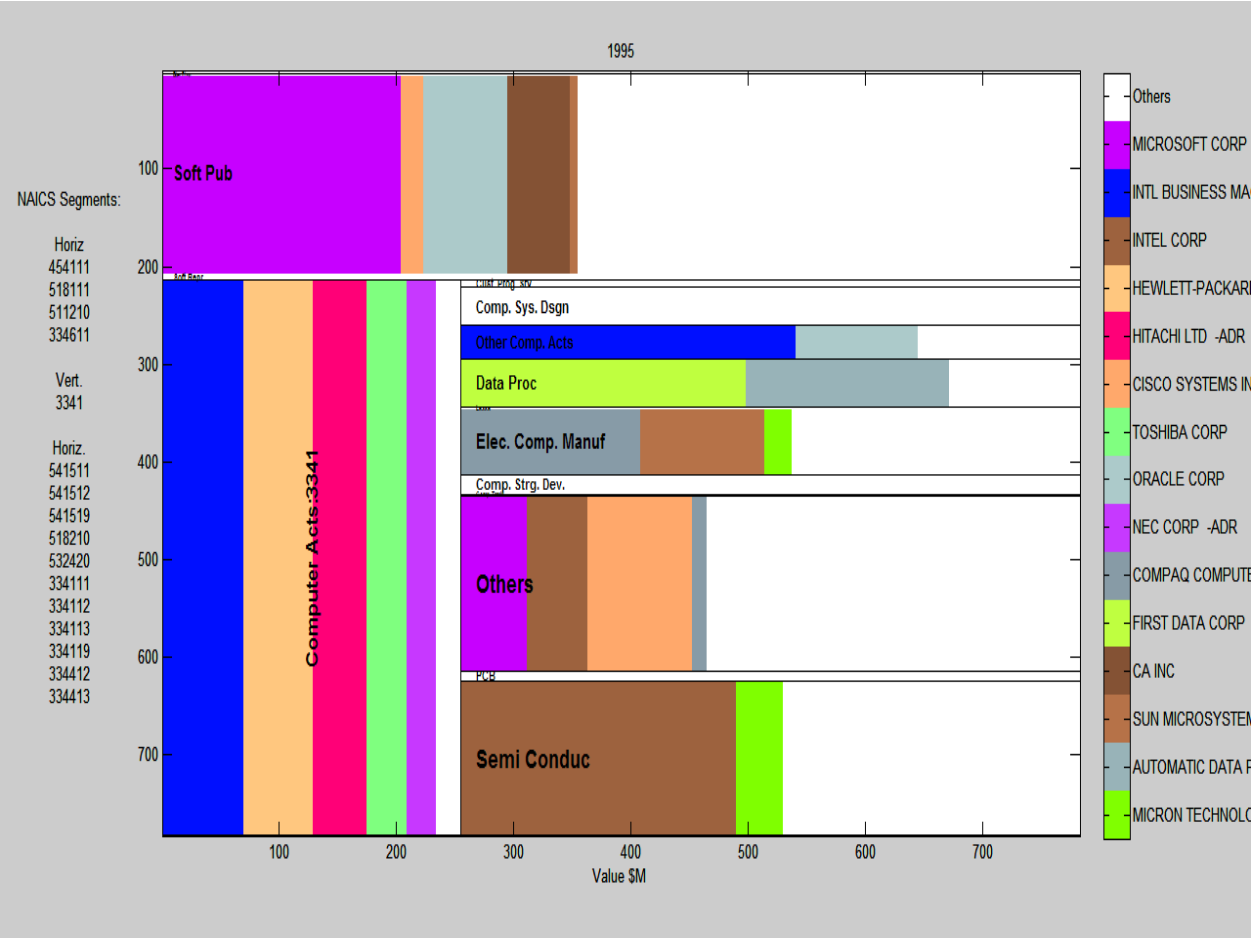
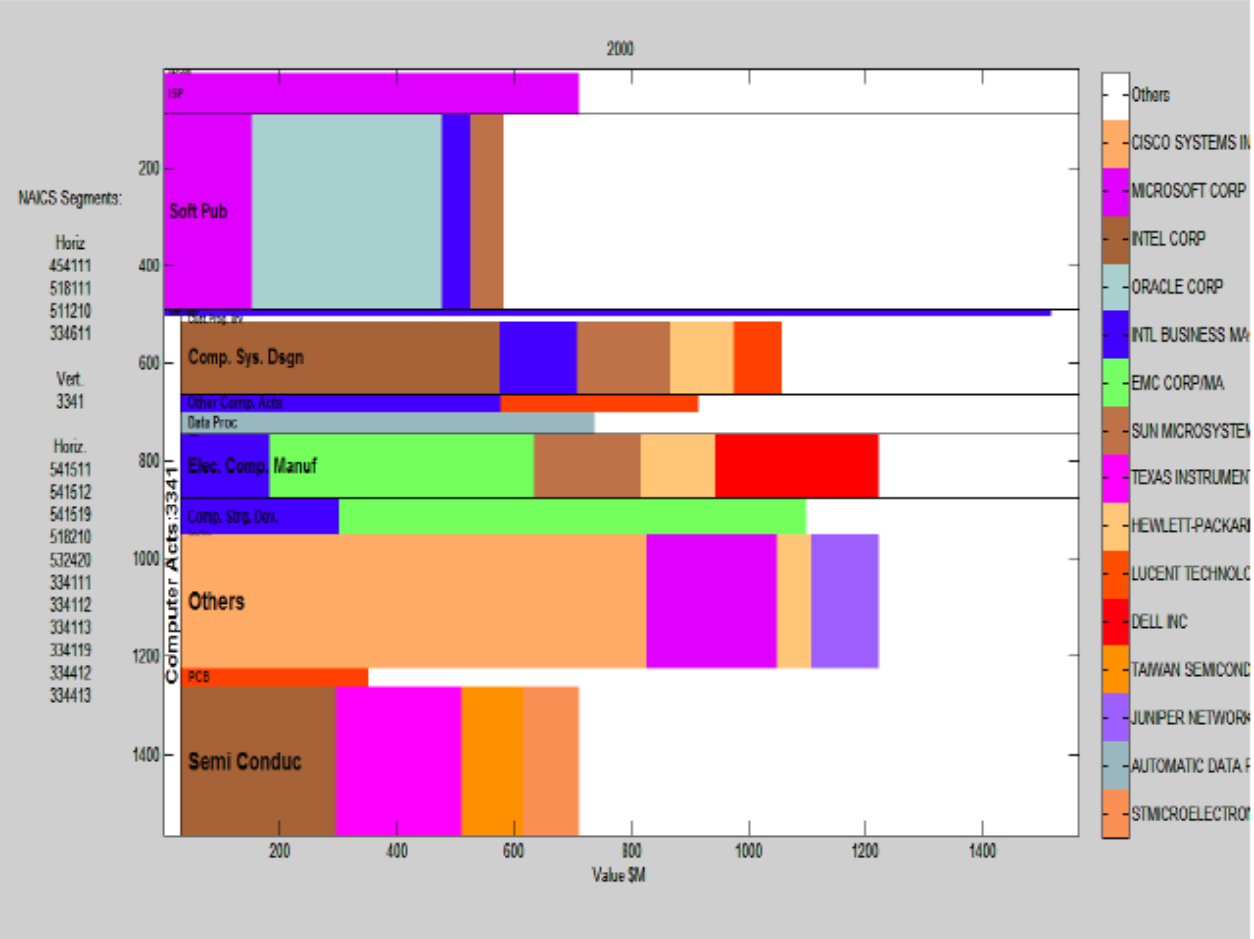


Plate 4 Distribution of Computer Industry Market Capitalization by Layer 2000



Three other important developments occurred toward the end of this period. First, as indicated, the arrival of the Internet provided the means to create computer-enabled "open exchange" platforms, including marketplaces, search engines and social media sites. By 2005 (not shown) the top tier of the industry included Google, Ebay, Yahoo, and a regenerated Apple. Second, the sponsors of the largest open platforms—the Internet, the Worldwide Web, and telecommunications platforms, were not single firms, but standards-setting organizations and (sometimes) governments. Increasingly, for-profit platform sponsors found themselves negotiating with representatives of their complementors, their users, and the public. Finally, in the late 1990s, loose-knit communities committed to transparent, non-proprietary designs emerged as creators and stewards of open source software codebases. Open source code began competing effectively with proprietary, closed source software, and for-profit firms had to adapt their strategies accordingly.

7.2 Conclusion—How Technology Shapes Organizations

Why did the transition to open platform systems in the computer industry begin in the 1980s and gather steam in the 1990s? During the 1970s, engineers throughout the industry, working in both software and hardware, became familiar with the concept of modularity and with modular architectures. The principles underlying the creation of a modular system were codified by Gordon Bell and Allen Newell and David Parnas in the early 1970s. Even within a single firm, the option value associated with modular systems was difficult to ignore. A modular architecture in turn is necessary for an open platform to achieve scale and diversity.

Firms (and other inventors) create modular architectures to increase the options available to the system architect and/or users. Opening up the modular system to suppliers and complementors further increases options and rates of innovation by introducing diversity and competition.

Also in the 1970s, the entry and survival of plug-compatible peripheral computer companies, attracted by the modularity of IBM System/360, created a ecosystem of hundreds of firms that specialized

in making modules for larger systems. The products of module makers were highly competitive and thus groups making the same components within vertically integrated firms found it hard to keep up with the price or performance of the "best of breed." The existence of these third parties in turn enabled entrepreneurial startups to become platform sponsors, competing with larger firms by setting up their own networks of suppliers and complementors.[82]

Last but not least, in the 1970s, control of the rate of technical progress in computers shifted from computer systems makers to semiconductor firms. Pure-play semiconductor firms like Intel, Texas Instruments, Advanced Micro Devices and National Semiconductor were in competition not only with each other but with state-subsidized Japanese companies. In late 1970s, U.S. firms were burned badly when they failed to keep up with Japanese firms in the transition to 64K memory chips. Following that episode, U.S. firms as a group became committed to the performance gains and price declines made explicit by Moore's Law and later codified in the National Semiconductor Roadmap. Given the advantages accruing to first movers, no semiconductor firm could afford to go slower than the "scheduled" pace.[83]

Because semiconductor chips were the fundamental building blocks of computer and communication systems, realized rates of change in semiconductor performance and prices basically set the pace for the rest of the industry. A systems maker like IBM might prefer to delay the introduction of the next generation computer. Introducing a new, superior model to the market when the older model is still viable is called "cannibalization." An incumbent always favors slow cannibalization, delaying the next generation until "the time is ripe."

However, when firms in the semiconductor industry came to believe in Moore's Law, and later

[82] The computer industry structure thus evolved according to the dynamic pattern described by Michael Jacobides in the mortgage banking industry (Jacobides, 2005). In our investigation of the mirroring hypothesis, Lyra Colfer and I found eight other instances in which industries split into horizontal layers after the introduction of a modular architecture (Colfer and Baldwin, 2016). Thus there appears to be a robust correlation between the invention of a modular technical architecture and a subsequent transition to horizontally layered industry structure.

[83] Flamm and Reiss (1993); Mollick (2006).

developed their National Roadmap, systems makers lost control over rates of change in their own systems. As IBM learned when it attempted to postpone introduction of Intel's 80386 processor, if the market leader did not keep pace with new semiconductor technology, other firms would happily bring the new technology to market in hopes of displacing the leader.

This combination of three factors—widespread knowledge about modular architectures; a growing ecosystem of specialized firms supplying modules; and consistent, rapid improvements in semiconductor performance and pricing—made it technically possible to create open supply networks and open networks of complementors, that is, open platform systems.

Yet it is safe to say that, in 1981, when IBM introduced PC, or even in 1985, the peak year, very little was known about how to manage or profit from an open platform system. This was uncharted territory. According to Grove, startup companies with limited resources would of necessity create open platforms. But a mature, successful company:

> would have its own semiconductor chip implementation, build its own computer around these chips according to its own design and in its own factories, develop its own operating system software … and market its own applications software…. This combination … would then be sold as a package by the company's own salespeople. This is what we mean by a vertical alignment. …
>
> The pluses were that … all parts would be made to work together as a seamless whole. … The minuses were that … [if] there was a problem, you couldn't throw out just one part of the vertical stack; you would have to abandon the entire stack, and that was a big deal.[84]

Note that the advantages of a vertical organization were precisely those of a step process with no bottlenecks: seamless integration with all parts efficiently working together. The disadvantages arose from lack of modularity: there were no options outside of the vertical stack. Businesses run this way also evolved towards stasis: "customers of the vertical computer companies tended to stay for a long time."[85]

Yet, from the early 1970s, virtually all new computers were designed as modular systems. Hence there was implicit tension between the traditional business model of vertical integration and the versatility

[84] Grove (1996) pp. 40-41.

[85] *Ibid.* p. 41.

and mix-and-match flexibility of the new technical architecture. It was only a matter of time before systems architects who perceived the value of a richer set of options, began to experiment with *open* modular platforms. These architects relied on the growing ecosystem of specialist firms to supply critical modules, implicitly settling for a smaller percentage of a larger pie.

Factors tilting the scales toward openness were: (1) heterogeneous user preferences, including the possibility of user innovation; (2) competition, especially the desire to overtake an incumbent or first mover; (3) resource constraints, both lack of capital and lack of capabilities to make some functional components.

However, firms embarking on this strategy had to find ways of protecting the modular system's visible information from becoming someone else's property. The IBM PC was one such experiment: it succeeded in establishing an open platform as the dominant architecture, but failed to capture value for IBM. In chapters below, I will discuss the experiences of Intel, Dell Computer and Sun Microsystems with open platforms. Intel exemplifies a forward-open platform; Dell represents a backward-open platform; and Sun pursued a strategy that was open in both directions. Following that, I will look at how the Internet gave rise to open exchange platforms and open source communities, leading to a new set of organizational forms and possibilities.

I have argued that the vertical-to-horizontal transition in the computer industry was an organizational response to a change in economic rewards to the competing technologies of rationalized step processes vs. open platform systems. The spread of modular architectures in conjunction with the rapid pace of change in semiconductor technology shifted the balance of rewards in this industry away from predictability and efficiency towards option value and mix-and-match flexibility. In this industry and others, managers must now support in platforms that encourage diversity and experimentation as well as coordinate complicated step processes. In this paper, I have set forth arguments that may indicate which technologiy is likely to dominate in a particular setting as well as when and where transitions may take place.

References

Abernathy, William J., Kim B. Clark and Alan M. Kantrow (1983) *Industrial Renaissance: Producing a Competitive Future for America*, New York: Basic Books.

Adner, Ron and Rahul Kapoor (2010) "Value Creation in Investment Ecosystems: How the Structure of Technological Interdependence Affects Firm Performance in New Technology Generations," *Strategic Management Journal* 31:306-333.

Ahlering, B. and Deakin, S., 2007. Labor regulation, corporate governance, and legal origin: A case of institutional complementarity?. *Law & Society Review*, *41*(4), pp.865-908.

Alchian, Armen A. and Harold Demsetz (1972) "Production, Information Costs, and Economic Organization," *American Economic Review* 52:777-795.

Atleson, J.B., 1983. *Values and assumptions in American labor law*. Boston, MA: University of Massachusetts Press.

Baker, George, Robert Gibbons, and Kevin J. Murphy. 1999. "Informal Authority in Organizations." *Journal of Law, Economics, and Organization*. 15,1:26-73.

Baldwin, Carliss Y. (2008) "Where Do Transactions Come From? Modularity, Transactions and the Boundaries of Firms," *Industrial and Corporate Change* 17(1):155-195.

Baldwin, Carliss Y. and Kim B. Clark (2000). *Design Rules, Volume 1, The Power of Modularity*, Cambridge, MA: MIT Press.

Berger, S., 2005. *How we compete: What companies around the world are doing to make it in today's global economy*. New York: Doubleday.

Berle, Adolph A. and Gardiner C. Means (1932; 1991) *The Modern Corporation and Private Property*, New York: Transactions Publishers.

Blair, M. M. (2003). Locking in capital: What corporate law achieved for business organizers in the nineteenth century. *UCLA Law Review*, *51*(2), 387-455.

Boudreau, Kevin J. and Andrei Hagiu (2011) "Platform Rules: Multi-Sided Platforms As Regulators," in *Platforms, Markets and Innovation*, (Annabelle Gawer, ed.) London: Edward Elgar.

Bresnahan, T.F. and Greenstein, S., 1999. Technological competition and the structure of the computer industry. *The Journal of Industrial Economics*, 47(1):1-40.

Brusoni, Stefano, Andrea Prencipe and Keith Pavitt (2001) "Knowledge Specialization, Organizational Coupling and the Boundaries of the Firm: Why Do Firms Know More Than They Make?" *Administrative Science Quarterly*, 46(4):597-621.

Colfer, L.J. and Baldwin, C.Y., 2016. The mirroring hypothesis: theory, evidence, and exceptions.

Industrial and Corporate Change, 25(5): 709-738.

Chafkin, M. and I. King (2016) "How Intel Makes a Chip," *BloombergBusinessweek* (June 9, 2016). Available at http://www.bloomberg.com/news/articles/2016-06-09/how-intel-makes-a-chip

Chandler, A. D. Jr. (1986). The beginnings of the modern industrial corporation. *Proceedings of the American Philosophical Society*, 130(4):382-389.

Chandler, Alfred D. (1962) *Strategy and Structure*, Cambridge, MA: MIT Press.

Chandler, Alfred D. (1977) *The Visible Hand: The Managerial Revolution in American Business*, Cambridge, MA: Harvard University Press.

Chandler, Alfred D. (1990) *Scale and Scope: The Dynamics of Industrial Capitalism*, Cambridge, MA: Harvard University Press.

Clark, Kim B. (1985) "The Interaction of Design Hierarchies and Market Concepts in Technological Evolution," *Research Policy* 14 (5): 235-51.

Coase, Ronald H. (1937) "The Nature of the Firm," *Economica,* 4(4):386-405, reprinted in *The Firm, the Market, and the Law,* University of Chicago Press, Chicago, IL, 33-55.

DeLamarter, R.T., (1986) *Big Blue: IBM's use and abuse of power*. Dodd, Mead & Company.

Drucker, P. F. (1993). *Concept of the Corporation*. Transaction Publishers, London.

Ferguson, Charles H. and Charles R. Morris (1993) *Computer Wars: How the West Can Win in a Post-IBM World,* New York, NY: Times Books.

Fields, Gary (2004) *Territories of Profit: Communications, Capitalist Development, and the Innovative Enterprises of G.F. Swift and Dell Computer*, Stanford CA: Stanford University Press.

Flamm, K. and Reiss, P.C., 1993. Semiconductor dependency and strategic trade policy. *Brookings Papers on Economic Activity. Microeconomics*, *1993*(1), pp.249-333.

Fransman, M. (undated) Industry Mapping: The Layer Model, Telecom Visions, http://www.telecomvisions.com/map/maptext3.php#Companies'%20Specialisation%20By

Freeland R. F. (2016) "The Employment Relation and Coase's Theory of the Firm," in *The Elgar Companion to Ronald H. Coase* (C. Menard and E. Bertrand, eds.) Northampton, MA: Edward Elgar.

Freeland, R.F. and Zuckerman, E. (2014) The problems and promises of hierarchy: A sociological theory of the firm. *Unpublished manuscript. University of Wisconsin and MIT Sloan School of Management*.

Frischmann, B.M. (2004) An economic theory of infrastructure and commons management. *Minnesota*

Law Review, 89:917-1030.

Frischmann, B.M., (2012) *Infrastructure: The social value of shared resources*. Oxford University Press.

Fukuyama, F., 1995. *Trust: The social virtues and the creation of prosperity*. New York: Free Press.

Galbraith, J.K., 1967. *The new industrial state*. Boston, MA: Houghton Mifflin

Garud, Raghu and Arun Kumaraswamy (1995) "Technological and Organizational Designs to Achieve Economies of Substitution," *Strategic Management Journal,* 17:63-76.

Gawer, A., 2014. Bridging differing perspectives on technological platforms: Toward an integrative framework. *Research Policy*, 43(7), pp.1239-1249.

Gawer, Annabelle and Michael A. Cusumano (2002) *Platform Leadership: How Intel, Microsoft and Cisco Drive Industry Innovation*, Boston, MA: Harvard Business School Press.

Grove, Andrew S. (1996). *Only the Paranoid Survive*, New York: Doubleday.

Halberstam, D. (1986). *The Reckoning*. New York: William Morrow.

Hansmann, Henry, Reinier H. Kraakman and Richard Squire (2006) "Law and the Rise of the Firm," *Harvard Law Review* 119(5):1335-1403.

Harry, M. J., and Schroeder, R. R. (2005). *Six Sigma: The breakthrough management strategy revolutionizing the world's top corporations.* Broadway Business.

Hilton, C. (1998) "Manufacturing Operations System Design and Analysis," *Intel Technology Journal Q4 1998*. Available at http://www.intel.com/content/dam/www/public/us/en/documents/research/1998-vol02-iss-3-intel-technology-journal.pdf

Hounshell, David A. (1985) *From the American System to Mass Production, 1800-1932*, Baltimore, MD: Johns Hopkins University Press.

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

Jacobides, Michael G. (2005). "Industry Change Through Vertical Dis-Integration: How and Why Markets Emerged in Mortgage Banking," *Academy of Management Journal*, 48(3):465-498.

Jacobides, M. G., Baldwin, C. Y. and Dizaji, R. (2007) "From the Structure of the Value Chain to the Strategic Dynamics of Industry Sectors," *Academy of Management Symposium Presentation,* Philadelphia, PA, August 7, 2007.

Jacobides, Michael G., Thorbjorn Knudsen and Mie Augier (2006) "Benefiting from Innovation: Value Creation, Value Appropriation and the Role of Industry Architecture," *Research Policy,* 35(8):1200-

1221.

Jacobides, M.G., Cennamo, C., and A. Gawer (2017) Towards a Theory of Business Ecosystems; unpublished manuscript.

Jaikumar, Ramachandran and Roger E. Bohn (1986) "The Development of Intelligent Systems for Industrial Use: A Conceptual Framework," in *Research on Technological Innovation, Management and Policy*, *Volume 3.* (R. S. Rosenbloom, ed.) Greenwich, CT: JAI Press.

Jensen, Michael C. and William H. Meckling (1976) "Theory of the Firm: Managerial Behavior, Agency Costs, and Ownership Structure," *Journal of Financial Economics,* 3(4):305-360, reprinted in Foundations of Organizational Strategy*,* Harvard University Press, Cambridge, MA.

Kaplan, A.D.H., 1964. *Big enterprise in a competitive system.* Washington DC: The Brookings Institution.

Kendall, H. P. (1911) "Types of Management: Unsystematized, Systematized, and Scientific," in *Dartmouth College Conferences, Addresses and Discussions of the Conference on Scientific Management Held October* (Vol. 12, No. 73, p. 14).

Landes, D. S. (1998). *Wealth and Poverty of Nations.* New York: W.W. Norton.

Landes, D.S., 1969. The Unbound Prometheus: Technological Change and Development in Western Europe from 1750 to the Present. Cambridge University Press.

Landes, David S. (1986) "What Do Bosses Really Do?" *The Journal of Economic History*, 46(3):585-623.

Langlois, Richard N. and Paul L. Robertson (1992). "Networks and Innovation in a Modular System: Lessons from the Microcomputer and Stereo Component Industries," *Research Policy,* 21(4): 297-313.

Litterer, J. A. (1963). Systematic management: Design for organizational recoupling in American manufacturing firms. *Business History Review*, *37*(04): 369-391.

Marx, L., 1994. The idea of "technology" and postmodern pessimism. In *Does Technology Drive History: The Dilemma of Technological Determinism,* (M.R. Smith and L. Marx, eds.) Cambridge MA: MIT Press.

Masten, S.E., 1988. A legal basis for the firm. *Journal of Law, Economics, & Organization*, *4*(1), pp.181-198.

Merton, Robert C. (1973) "Theory of Rational Option Pricing," *Bell Journal of Economics and Management Science*, 4(Spring): 141-183; reprinted in *Continuous Time Finance*, Basil Blackwell, Oxford, UK, 1990.

Milgrom, Paul and John Roberts (1990) "The Economics of Manufacturing: Technology, Strategy and Organization," *American Economic Review* 80 (3): 511-28.

Milgrom, P., & Roberts, J. (1995). Complementarities and fit strategy, structure, and organizational change in manufacturing. *Journal of accounting and economics*, *19*(2), 179-208.

Misa, T.J., 1998. *A nation of steel: The making of modern America, 1865-1925*. Baltimore MD: Johns Hopkins University Press.

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

Moore, G. E. (1965) "Cramming more components onto integrated circuit"s, *Electronics* 38(8):114 ff.

Moore, G. E. (1975) "Progress in digital integrated electronics". *IEEE Solid-State Circuits Society Newsletter*, *20*(3):11-13.

Moore, J. F. (1996). *The Death of Competition: Leadership & Strategy in the Age of Business Ecosystems*. New York: HarperBusiness

Navin, T.R., 1970. The 500 largest American industrials in 1917. *Business History Review*, *44*(03), pp.360-386.

Navin, T.R. and Sears, M.V., 1955. The rise of a market for industrial securities, 1887-1902. *Business History Review*, *29*(02), pp.105-138.

Nelson, D. (1974). Scientific management, systematic management, and labor, 1880–1915. *Business History Review*, *48*(04), 479-500.

Noble, David F. (1984) *Forces of Production: A Social History of Industrial Automation*, Oxford: Oxford University Press.

Parker, G.G., Van Alstyne, M.W. and Choudary, S.P. 2016. *Platform revolution: How networked markets are transforming the economy--and how to make them work for you*. New York: W.W. Norton & Company.

Perrow, Charles (1986) *Complex Organizations: A Critical Essay*, New York: McGraw Hill.

Prencipe, Andrea, Andrew Davies and Mike Hobday, eds. (2003) *The Business of Systems Integration*, Oxford, UK: Oxford University Press.

Pugh, Emerson W., Lyle R. Johnson, and John H. Palmer (1991) *IBM's 360 and Early 370 Systems,* Cambridge, MA: MIT Press.

Rochet, Jean-Charles and Jean Tirole (2003) "Platform Competition in Two-sided Markets," *Journal of the European Economic Association*, 1(4): 990-1029.

Roe, M.J., 1991. A political theory of American corporate finance. *Columbia Law Review*, *91*(1), pp.10-67.

Roe, M.J., 1996. *Strong managers, weak owners: The political roots of American corporate finance*.

Princeton University Press.

Rosenberg, N., & Birdzell, LE Jr. (2008). *How the West grew rich: The economic transformation of the industrial world.* Basic Books, New York.

Rothschild, M., & Stiglitz, J. E. (1970). Increasing risk: I. A definition. *Journal of Economic theory*, *2*(3), 225-243.

Roy, W.G., 1999. *Socializing capital: The rise of the large industrial corporation in America*. Princeton, NJ: Princeton University Press.

Sanchez, Ronald A. and Joseph T. Mahoney (1996). "Modularity, Flexibility and Knowledge Management in Product and Organizational Design," *Strategic Management Journal*, 17: 63-76; .

Schilling, Melissa A. (2000). "Toward a General Systems Theory and its Application to Interfirm Product Modularity," *Academy of Management Review* 25(2):312-334.

Schumpeter, Joseph A. (1942). *Capitalism, Socialism, and Democracy*, New York: Harper & Brothers.

Servan-Schreiber, J.J., 1968. *The American Challenge*. New York: Atheneum.

Shenhav, Y., 1995. From chaos to systems: The engineering foundations of organization theory, 1879-1932. *Administrative Science Quarterly*, pp.557-585.

Simon, H.A., 1990. A mechanism for social selection and successful altruism. *Science*, *250*(4988), pp.1665-1669.

Simon, Herbert A. and Mie Augier (2002) "Commentary on 'The Architecture of Complexity'," in *Managing in the Modular Age: Architectures, Networks, and Organizations,* (R. Garud, A. Kumaraswamy, and R. N. Langlois, eds.) Blackwell, Oxford/Malden, MA.

Smith, Adam (1994) *An Inquiry into the Nature and Causes of the Wealth of Nations*, (E. Cannan, ed.) New York, NY: Modern Library.

Sturgeon, Timothy J. (2002). "Modular Production Networks: A New American Model of Industrial Organization," *Industrial and Corporate Change*, 11(3): 451-496.

Thompson, James D. (1967) *Organizations in Action: Social Science Bases of Administrative Theory*, New York, NY: McGraw-Hill.