

**POWER MANAGEMENT AS A SYSTEM-LEVEL INHIBITOR OF
MODULARITY IN THE MOBILE COMPUTER INDUSTRY**

by

Samuel K. Weinstein

Bachelor of Science in Electrical Engineering
The Ohio State University, 1996

Submitted to the System Design & Management Program in
Partial Fulfillment of the Requirements for the Degree of

Master of Science in Engineering and Management

at the

Massachusetts Institute of Technology

June 2004

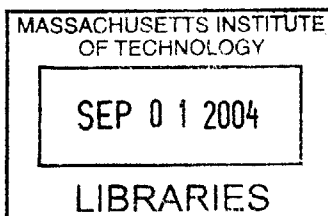
©2004 Samuel K. Weinstein. All rights reserved. The author hereby grants to MIT permission to reproduce, and to distribute publicly, paper and electronic copies of this thesis document in whole or in part.

Signature of Author _____
Samuel K. Weinstein
System Design & Management Program

Certified by _____
Daniel E. Whitney
Thesis Supervisor
Senior Research Scientist, Center for Technology, Policy, and Industrial Development

Accepted by _____
Tomas J. Allen
Co-Director, LFM/SDM
Howard W. Johnson Professor of Management

Accepted by _____
David Simchi-Levi
Co-Director, LFM/SDM
Professor of Engineering Systems



BARKER

POWER MANAGEMENT AS A SYSTEM-LEVEL INHIBITOR OF MODULARITY IN THE MOBILE COMPUTER INDUSTRY

by Samuel K. Weinstein

Submitted to the System Design & Management Program in
Partial Fulfillment of the Requirements for the Degree of
Master of Science in Engineering and Management

Abstract

Since the mid-90s, the computer industry has been very modular with respect to both product architecture and industry structure. The growing market size of mobile computers means that the challenges facing this segment are starting to affect the direction of the industry. It is argued in this paper that power management in mobile computers is forcing the industry in the direction of more integral product solutions and, hence, a more integral industry structure. That is to say, the industry is assuming a structure similar to the early days of mainframe computers when one firm delivered the entire proprietary integral system. Furthermore, this trend towards more integrality in mobile computer systems is due to fundamental physical attributes of the system; specifically, that information transfer systems lend themselves more readily to modular architectures than systems that transfer significant power. Thus, as processors and mobile computers become more powerful, they start to behave more like power transfer systems and side effects of this power, such as heat, require a more integral approach to managing it. A “free body” diagram framework is presented which provides a way of thinking about how integrality forces are acting on an industry’s trajectory. Evidence is presented showing how the dominant player in the computer supply chain, Intel, is exhibiting this vertical/integral behavior in a number of ways.

Thesis Advisor: Daniel E. Whitney
Title: Senior Research Scientist

THIS PAGE INTENTIONALLY LEFT BLANK

Table of Contents

Abstract	3
List of Figures	6
List of Tables	7
Acknowledgements	8
1 Introduction	10
1.1 A Systems Approach.....	10
1.2 Thesis Organization	12
2 Background and Literature	14
2.1 Modularity and Integrality	14
2.1.1 Definitions of Modularity and Integrality.....	15
2.1.2 Arguments For / Against Modularity.....	17
2.1.3 Definitions of Types of Modularity for This Thesis.....	20
2.2 Information and Power Systems	23
2.3 The Fine / Whitney Framework for the Dynamic Modularity of Industries.....	26
2.4 Moore’s Law: Cultural Icon, Industry Driver, and Corporate Monument	30
2.4.1 The Original Law	30
2.4.2 The Real Law	31
2.4.3 The Self-Fulfilling Prophecy	33
2.4.4 The Method Behind the Madness	34
3 The Mobile Computer Industry	37
3.1 A Brief History of the Computer Industry’s Structure	37
3.2 Along Comes the Mobile.....	41
3.3 A Short Reprise Before the Punch Line.....	44
4 Power as a Chief Limiter in Mobile Computer Modularity	45
4.1 Pushing Moore’s Law: CMOS Scaling and Power.....	47
4.2 Peak Power Density and Cooling Costs.....	52
4.3 Market Pull: Average Power and Battery Life	57
5 A Framework for the Dynamic Modularity of Industries	63
5.1 The Framework.....	63
5.2 Transaction Costs.....	65
5.3 The Four Quadrants	66
5.4 Application of the Framework to the Computer Industry.....	69
6 Signs of Verticality	72
7 Conclusion	81
Bibliography	88

List of Figures

Figure 1: Example of a Modular and Integral System.....	15
Figure 2: Modular vs. Integral Architecture	16
Figure 3: Modularity in Axiomatic Design.....	20
Figure 4: The Fine / Whitney "Double Helix"	28
Figure 5: Circuit Density of Intel Microprocessors 1970-2000.....	31
Figure 6: Speed of Intel Microprocessors 1970-2000.....	32
Figure 7: Computer Industry Structure Prior to 1985.....	39
Figure 8: Computer Industry Structure After 1985.....	40
Figure 9: The 1979 Grid Compass and the 1989 NEC UltraLite	41
Figure 10: Causal Loop Diagram of Traditional Scaling's Effect on Power	49
Figure 11: Emergence of Standby Power as Significant Design Parameter.....	51
Figure 12: Causal Loop Diagram of Cooling Cost and Size Constraints for a Laptop	54
Figure 13: Processor Power Limits for Present Cooling Technology	55
Figure 14: Causal Loop Diagram of Laptop Speed / Weight / Size Tradeoff	57
Figure 15: Processor Power Limits for Present Mobile Supply Technology	59
Figure 16: Modularity Framework.....	64
Figure 17: Applied Modularity Framework - "Free Body Diagram"	70
Figure 18: Current Industry Structure (Horizontal Going to Vertical).....	72
Figure 19: Intel's "Holistic" Approach to Power Management	75
Figure 20: Intel Mobile Internet PC 2004 Platform Vision	78
Figure 21: The Gateway Media Center Desktop Designed by Intel.....	79

List of Tables

Table 1: Pros and Cons of Modularity and Integrality	18
Table 2: Examples of Signal Processors and Power Processors.....	24
Table 3: The Effect of Modularity on Outsourcing Decision.....	27
Table 4: Forces in the Double Helix	29
Table 5: Recent Approximate CMOS Scaling Factors.....	50
Table 6: Options for Mobile Computer Power Management Solutions	62
Table 7: Investments in Power Management Technology by Intel Capital.....	77

Acknowledgements

First and foremost, I would like to thank my wife, Caitlin, and my parents, Fred and Mary Weinstein. I never would have had the dream of attending MIT without my parents' encouragement throughout my life and I never would have had the opportunity to do so without my wife's love and support throughout the last year of school.

This is especially true in light of all the transitions my wife and I have gone through in the last year from moving to a 500 square foot efficiency apartment in Cambridge and her changing jobs to preparing for the birth of our first child in May. It has been an exhausting but exciting year and now it is time to reap the rewards of our sacrifices.

I would also like to thank Dan Whitney, my advisor, for his uncanny ability to find a common ground between my interests and his so that both of us remained engaged in this thesis process, for his continued ability to find time in his busy schedule to discuss my research and assuage my fears (that most everyone ever writing a thesis has had) that I was not making progress, and for his nudges in the right direction without being so descriptive in his guidance as to stifle my creativity and limit my thinking. Dan is an engineer first, but truly a great systems thinker—ESD is fortunate to have him as a faculty member.

Lastly, I want to acknowledge the part of my degree that is sometimes taken for granted: my SDM classmates. This past year wouldn't have been worthwhile without all of the valuable shared experiences in class discussion and collaboration on assignments by the diverse and talented members of SDM03. Furthermore, I wouldn't have made it through the year without their support and commiseration. All of the afternoons spent at the Muddy Charles were not time wasted that could have been spent doing coursework—they were time well invested in forming friendships to last a lifetime.

The SDM Program has without a doubt been the most difficult thing I have ever tried to do (although I'm sure that will change when I become a parent in a few months), however, I wouldn't trade the experiences of the last year for anything.

I extend my sincerest gratitude to you all. Thank you.

THIS PAGE INTENTIONALLY LEFT BLANK

1 Introduction

1.1 A Systems Approach

This thesis attempts to explain changes taking place in the mobile computer industry by examining dynamic forces acting on it in the context of a framework established herein.

The treatment of these technology- and market-driven forces, as well the theoretical basis for the framework that is used to conceptualize the consequences of the forces, are drawn from the literature. As with most systems thinking approaches, the goal here is to “get one’s mind around” a complex phenomenon with many layers and nuances by

- breaking down the phenomenon into its fundamental component parts
- providing a tool to help strip away some of the complexity through visualizations and links that the human mind can more easily comprehend and manipulate.

Specifically, this systems approach involves the description of the driving forces and the subsequent development of a framework that attempts to link changes in modularity of design to changes in the organization of firms in the mobile computer industry.

Throughout this thesis, the discussion will focus primarily on a single firm: Intel Corporation (Nasdaq: INTC). Thus, the approach is to look at firms in the mobile computer industry, primarily through analysis of the forces acting on Intel’s strategy, development efforts, and investments. This simplification is possible because:

- Intel has over 80% market share in desktop and mobile PC microprocessors
- Intel is a tier one supplier of the module that controls the value chain (since the microprocessor is the most important value-add component)

Thus, the forces acting on Intel are a good approximation of the forces acting on the mobile computer industry because they control the value chain. So, to a large extent, the direction Intel takes is necessarily the direction the mobile computer industry takes.

1.2 Thesis Organization

The organization of this thesis is as follows:

Chapter 2 describes the literature and gives some background information. First, the concept of modularity is introduced and defined for the context of the argument presented herein regarding product architecture and industry structure. Second, the fundamental differences between information and power systems and the direct impact on achievable levels of modularity for each type of system are discussed. This is followed by a discussion of the “double helix” of industry organization, which deals with the cyclical nature of product architecture between integral and modular and industry structure between vertical and horizontal. Lastly in Chapter 2, Moore’s Law and its implications for the computer industry and society at large are reviewed.

Next, a brief history of the computer industry’s structure is given in Chapter 3, including the major turning points in the transition from vertical to horizontal. This is followed by the history of the mobile computing trend and a description of the challenges facing this small form factor version of the PC. Chapter 4 continues the discussion started in the previous chapter with a detailed description of some of the technical challenges that the industry is presently facing with regard to power management. It is seen that continued traditional electronic circuit scaling will lead to chip peak power densities that cannot cost-effectively be cooled in a laptop and average system power ratings that cannot be powered by existing battery technology for a meaningful amount of time.

The following chapter, number 5, presents a framework for thinking about how the forces acting on an industry will affect the modularity of the product architecture and the trajectory of the industry going forward. This includes introducing the important effect of transaction costs. The framework is then applied to the computer industry in the context of the forces created by the solutions to the power management problems mentioned in the previous chapters. It is seen that the computer industry is moving from a “modular cluster” organization to one with a more integral product architecture and a more integral industry structure.

Chapter 6 provides recent evidence that the mobile computer industry is, in fact, moving along the trajectory predicted by the framework. Examples are cited from Intel and their numerous vertical expansions, many of which are a direct result of the aforementioned power management issues. The final chapter summarizes the arguments of the previous chapters and suggests areas for future research by posing some unanswered questions that this thesis raises.

2 Background and Literature

2.1 Modularity and Integrality

Fundamental to the way development projects, firms, and even entire industries are organized is the architecture of the products they are designing and producing. A simple definition of architecture is the way something is put together. More specifically, it is all the components of a system and their interconnections.¹ Modularity, then, is defined as a property of the system that describes how closely the elements of form map to the elements of function. An architecture that is more modular has elements of form that are closely matched to the elements of function. For example, most software programs (see Figure 1) are highly modular with each subroutine dedicated to a specific function. An architecture that is less modular, or more integral, has elements of form that are not closely matched to the elements of function. A bridge is one such system with all structural members working together to deliver the load bearing function and many of the members working together to deliver the path for transportation function. The following sections below describe these definitions in greater detail.

¹ [Crawley, 2003]

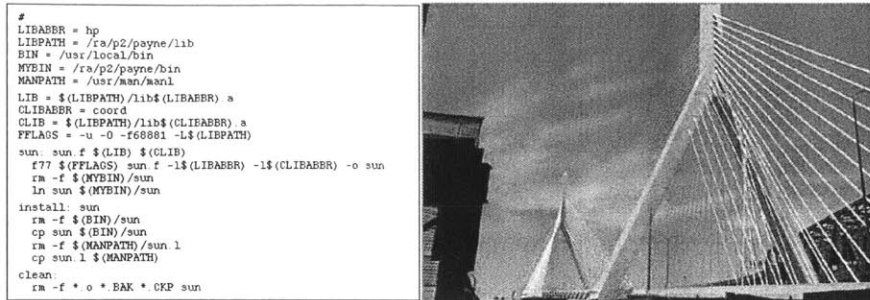


Figure 1: Example of a Modular (FORTRAN code) and Integral (Zakim Bridge in Boston) System

2.1.1 Definitions of Modularity and Integrality

Ulrich and Eppinger describe a modular architecture as one in which “each physical chunk² implements a specific set of functional elements and has well-defined interactions with other chunks.” This definition can be used to inspect every architecture that has physical parts, but others, such as software architectures, are perhaps better suited by removing the “physical” adjective. Thus, a completely modular architecture is one in which each functional element of the system is accomplished by exactly one structural element, or chunk. Conversely, integrality in architecture requires that the system have at least one of the following:

- Functional elements implemented using more than one chunk
- A single chunk which implements many functions

² “Chunks” are the major structural elements of a system that together accomplish the product’s main function(s)

- Ill-defined interactions between chunks.³

Figure 2 below illustrates this distinction. Notice that in a completely modular architecture, there is a clean 1-to-1 mapping of the functions to the chunks all the way down through the hierarchy. That is to say, Sub-function 1 is implemented by Large Chunk 1, Individual Function 4 is implemented by Small Chunk 4 and so on. However, in an integral architecture, this is not the case. For example, Individual Function 4 is implemented by Small Chunk 3 and Large Chunk 2. Furthermore, Small Chunk 3 also helps implement Individual Functions 1 and 2, while Large Chunk 2 serves the additional purposes of implementing Individual Functions 2 and 5.

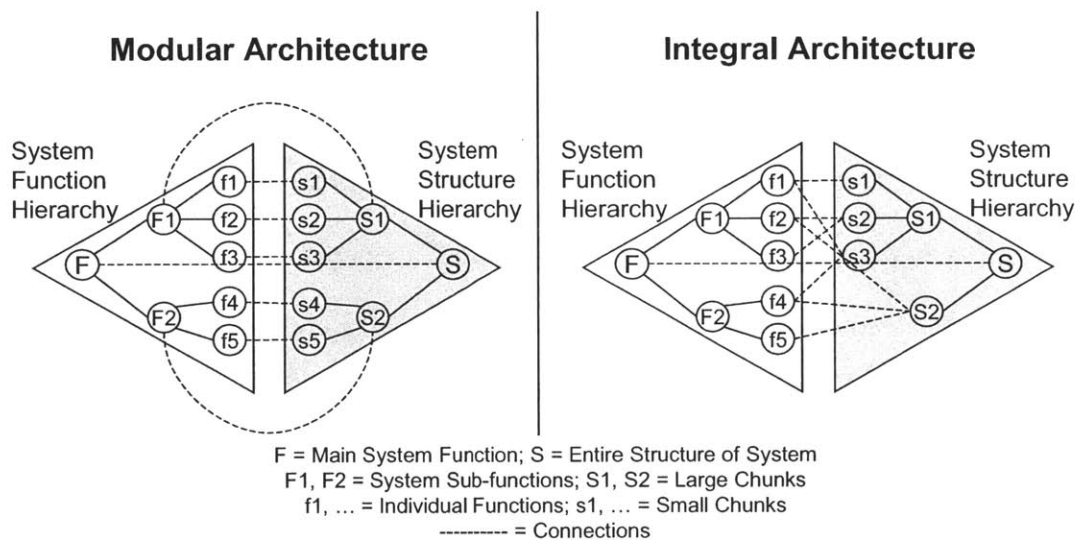


Figure 2: Modular vs. Integral Architecture⁴

³ [Ulrich and Eppinger, 2000]

⁴ Adaptation of a figure from [Fujimoto, 2002]

Baldwin and Clark define modularity in terms of interdependence within and independence among modules plus management of complexity by information hiding within each module.⁵ Thus, the design becomes a collection of abstractions and interfaces (modules and connections) that can be designed independently by describing a set of design rules.

Clearly, a modular architecture is more easily understandable or at least decomposable into less complex parts. However, the next sections will illustrate that a modular system architecture is not always the best choice. Indeed, there are fundamental limitations to some systems that prohibit a conscious choice altogether.

2.1.2 Arguments For / Against Modularity

The laundry list of pros and cons in modularity of architecture is best summarized in a table format, shown below.

For Modularity / Against Integrality	Against Modularity / For Integrality
Product variety and change (substantial reduction in the fixed costs of introducing architecture variants)	Product sophistication, integrity, and higher performance
Must use for “open” architecture products, but can also be used for “closed”	Can help protect intellectual property for “closed” architecture products

⁵ [Baldwin and Clark, 2000]

Must use for a delayed differentiation strategy	Products may not be competitive if too many compromises are made to make the platform ⁶ module(s) fit the product
Platform components can benefit from economies of scale since production volume for these components is the total of the production volume of the products in which the platform components are used	Modular platforms may have longer development time since requirements for many products are considered, as opposed to a single set of product requirements
Design reuse can shorten development time and for evolutionary (as opposed to revolutionary) products	Designers have a hard time with reuse and tend to de-value “old” ideas, so they can do something “creative” and “novel”
Axiomatic Design (see below)	Modularity implies the use of “design rules” in the design process; this adds extra tasks to the development effort since the design must be checked against these rules
More reuse leads to more focus on the overall design, which translates into better architecture, tighter integration of components, and lower unit costs	For a “white space” or revolutionary design, modularity can result in too great a focus on sub-system improvements, rather than system-level innovation
Re-use of complex components designed previously to make an even more complex “meta-system” of systems that would have been too complex to build from scratch	The upfront costs of developing a platform are substantial and the market may not bear the resulting cost; related to this point is that low-end variants of the architecture are typically over-designed and are more expensive to produce on a per unit basis
Making money on products that are near commodities may require differentiation through offering many variants	Many variants of a product, some of which have a low profit margin, can cannibalize a firm’s sales

Some ideas in the table from these references:^{7 8 9 10}

Table 1: Pros and Cons of Modularity and Integrality

⁶ Platform modules are hardware or software components that are reused in more than one product.

⁷ [Robertson and Ulrich, 1998]

⁸ [Meyer and Lehnerd, 1997]

⁹ [Dalziel, 2002]

¹⁰ [Krishnan and Gupta, 2001]

A more “scientific” argument in favor of modularity is that given by Suh¹¹ with his concept of axiomatic design. This is a decomposition process going from customer needs to functional requirements (FRs), to design parameters (DPs), and then to process variables (PVs), thereby crossing the four domains of the design world: customer, functional, physical, and process. Axiomatic design is based upon two basic axioms. The first is that FRs must be independent of each other. The goal of this exercise is to identify DPs so that each FR can be satisfied without affecting the other FRs. The second axiom is that the information content of the design must be minimized. That’s “information” as in the measure of one’s freedom of choice, the measure of uncertainty, which is the basis of information theory. Information content is the logarithm of the inverse of the probability of delivering the FR.

In the diagram below, functional requirements are mapped to design parameters in matrix form. The diagram shows that in order for a design to be decoupled, this matrix must be triangular and to be completely uncoupled, it must be diagonal.

¹¹ [Suh, 1990]

$$\begin{array}{ccc}
 \left. \begin{array}{l} \{FR11\} \\ \{FR21\} \\ \{FR22\} \\ \{FR31\} \end{array} \right\} = \begin{array}{cccc} [X & 0 & 0 & 0] \\ 0 & X & 0 & 0 \\ 0 & 0 & X & 0 \\ 0 & 0 & 0 & X \end{array} \left. \begin{array}{l} \{DP11\} \\ \{DP21\} \\ \{DP22\} \\ \{DP31\} \end{array} \right\} & \begin{array}{l} \text{Decoupled} \\ \\ \\ \text{Coupled} \end{array} & \begin{array}{cccc} [X & 0 & 0 & 0] \\ 0 & X & 0 & 0 \\ 0 & X & X & 0 \\ X & 0 & X & X \\ [X & X & 0 & X] \\ 0 & X & 0 & X \\ X & 0 & X & 0 \\ 0 & X & X & X \end{array} \\
 \text{Uncoupled} & &
 \end{array}$$

Figure 3: Modularity in Axiomatic Design

Baldwin and Clark tout the power of modularity in their seminal work. In fact, their book focuses on how modularity has affected the computer industry. The effect of modularization on this industry has been so positive that “today, this highly evolved, decentralized social framework allows tens of thousands of computer designers to work on complex, interrelated systems in a decentralized, yet highly coordinated way.”¹² This thesis identifies trends that make this decentralization harder to maintain.

2.1.3 Definitions of Types of Modularity for This Thesis

This research defines two types of modularity that are important to the discussion. There is *product architecture modularity*, which is the degree to which a product can be broken into components that can be individually designed and manufactured according to

¹² [Baldwin and Clark, 2000]

predetermined design rules; and there is *industry structure modularity*, which is the degree to which those individual components are designed and manufactured by different firms. Thus, the terms “modular industry structure” and “horizontal industry structure” will be used interchangeably throughout this thesis, as will “integral industry structure” and “vertical industry structure,” or “vertical integration.”

Equally useful for the following discussion is this definition of modularity: modular systems have the property that all important interactions occur across defined interfaces, while integral systems are those in which significant interactions between chunks occur across both defined interfaces and undefined interfaces, sometimes called "sneak paths."¹³ It will be seen that this is especially true as the discussion of information and power systems begins in the following section.

It should also be pointed out that integrality has a different meaning than what most microelectronics industry insiders refer to as integration. Integration means putting more functions on a single die, or computer chip, or packaging more chips together so that there is less slow down in propagating the signals off-chip and onto another one within the computer system. The trade-off is between the amount of performance benefit the architect can obtain through integrating more functions and the cost of the integration in either yield (a larger die will have more probability of being scrapped due to defects) or

¹³ Discussion with Dan Whitney, January 16, 2004.

testing costs (placing more chips in a single package and doing so reliably). While this can be considered one form of integrality, it only addresses the chip level. When dealing with the entire computer system, there are many other architecture decisions that make the system more integral beyond simply integrating more functions on a chip. This will become evident in the discussion of power management below.

2.2 Information and Power Systems

Many have written on the topic of the inherent differences between mechanical and information systems. Whitney¹⁴ extensively explores this topic in the context of integrality versus modularity of design. Information systems', such as software or digital hardware (signal processors), value-related function is information transfer; complex electro-mechanical systems' (CEMs) main function is significant power transfer (see Table 2). This fundamental difference in system function has significant consequences for modularity of architecture. Specifically, signal processors lend themselves to almost complete modularity of design and all of the design automation efficiencies (component libraries, design rule checking) that go along with it, while mechanical systems are "stuck" with integral designs that require complex interactions of subsystems and components and, therefore, significant redesign and consideration of system-level side effects when part of the design is changed. Table 2 shows some examples of each of these types of systems and their associated peak powers.

¹⁴ [Whitney, 1996]

SIGNAL PROCESSORS	PROCESS AND TRANSMIT SIGNIFICANT POWER
four digit mechanical gear gas meter dial (1 mW?)	Polaroid camera (30W peak?)
ball-head typewriter (30 mW peak at the ballhead?)	missile seeker head(50W peak?)
sewing machine (1 W?)	laser printer (1 KW, much of which is heat)
Marchand calculator (10W?)	automobile automatic transmission (50 KW+)
	automobile (100 KW+)(half or more dissipated as heat from engine)
	airplane (10 MW±)
	ship (40 MW+)

Table 2: Examples of Signal Processors and Power Processors¹⁵

One of the main reasons for this difference is the impedance of the connections between components. VLSI circuits and software component modules can easily be designed so that there is significant impedance mismatching and, hence, no “backloading” occurs. In other words, since the information can be transferred at low power, the output of one module is not significantly affected by the input of the next module(s) in an undesirable way. Thus, information systems can be verified to a large extent at the component level because the components will behave in a predictable way when they are assembled.

Conversely, systems that transfer significant power cannot be cost-effectively designed to mismatch impedances to the point where system-level side effects become unimportant.

¹⁵ From [Whitney, 1996]

The classic analogy is that in order to obtain such a ratio of input to output impedance, for a CEM (in this case, a jet engine), “the turbine would be the size of a house and the propeller the size of a muffin fan. No one will build such a system. Instead, mechanical system designers must always match impedances and accept backloading. This need to match is essentially a statement that the elements cannot be designed independently of each other.”¹⁶ In other words, there is necessarily a limit to the amount of modularity of architecture due the backloading in CEMs.

This limit to the modularity of systems that transmit significant power has profound consequences for the future design of mobile computers and, indeed, the entire computer industry. As will be seen in the following chapters, not only does the relentless pursuit of Moore’s Law require more and more power from computer systems, but also meeting the demands of mobile computing customers requires better system-level power management solutions. Thus, the limitations that affect the modularity of CEMs will come to bear on mobile computers because they will *become* CEMs.

¹⁶ [Whitney, 1996]

2.3 The Fine / Whitney Framework for the Dynamic Modularity of Industries

One of the most important aspects of systems thinking is the dynamic nature of both the internal components and connections comprising the system and external forces acting on the system. An interesting framework for the dynamic nature of the modularity or integrality of product architectures and the subsequent horizontal or vertical organizations of firms in the industry is presented by Fine and Whitney.¹⁷

Their argument follows this reasoning:

Outsourcing creates two different kinds of dependency: least risky is the dependency for capacity (the firm retains the knowledge); the most risky is the dependency for that knowledge. However, the degree of risk in either of these two cases is influenced by the degree to which the architecture is modular. This is summarized Table 3 below:

¹⁷ [Fine and Whitney, 1996]

MATRIX OF DEPENDENCY AND OUTSOURCING

		DEPENDENT FOR KNOWLEDGE	DEPENDENT FOR CAPACITY
OUTSOURCED ITEM IS DECOMPOSABLE	<p>A POTENTIAL OUTSOURCING TRAP</p> <p>YOUR PARTNERS COULD SUPPLANT YOU. THEY HAVE AS MUCH OR MORE KNOWLEDGE AND CAN OBTAIN THE SAME ELEMENTS YOU CAN.</p>	<p>BEST OUTSOURCING OPPORTUNITY</p> <p>YOU UNDERSTAND IT, YOU CAN PLUG IT INTO YOUR PROCESS OR PRODUCT, AND IT PROBABLY CAN BE OBTAINED FROM SEVERAL SOURCES. IT PROBABLY DOES NOT REPRESENT COMPETITIVE ADVANTAGE IN AND OF ITSELF. BUYING IT MEANS YOU SAVE ATTENTION TO PUT INTO AREAS WHERE YOU HAVE COMPETITIVE ADVANTAGE, SUCH AS INTEGRATING OTHER THINGS</p>	
OUTSOURCED ITEM IS INTEGRAL	<p>WORST OUTSOURCING SITUATION</p> <p>YOU DON'T UNDERSTAND WHAT YOU ARE BUYING OR HOW TO INTEGRATE IT. THE RESULT COULD BE FAILURE SINCE YOU WILL SPEND SO MUCH TIME ON REWORK OR RETHINKING.</p>	<p>CAN LIVE WITH OUTSOURCING</p> <p>YOU KNOW HOW TO INTEGRATE THE ITEM SO YOU MAY RETAIN COMPETITIVE ADVANTAGE EVEN IF OTHERS HAVE ACCESS TO THE SAME ITEM.</p>	

Table 3: The Effect of Modularity on Outsourcing Decision¹⁸

The work goes on to show that the structure of the product and that of the industry can be quite similar; both may tend to be integral/vertical or modular/horizontal at any one time. More importantly, these configurations appear to be unstable for a variety of related technical and economic reasons, and have been found in several industries to cycle from

¹⁸ From [Fine and Whitney, 1996]

one form to the other and back. The bottom line is that no outsourcing or architecture modularity policy that seems suitable at a given time is likely to remain suitable, due to such upsetting factors as technological advances, regulatory changes, and other economic shifts. This is summarized in the “double helix”¹⁹ model below:

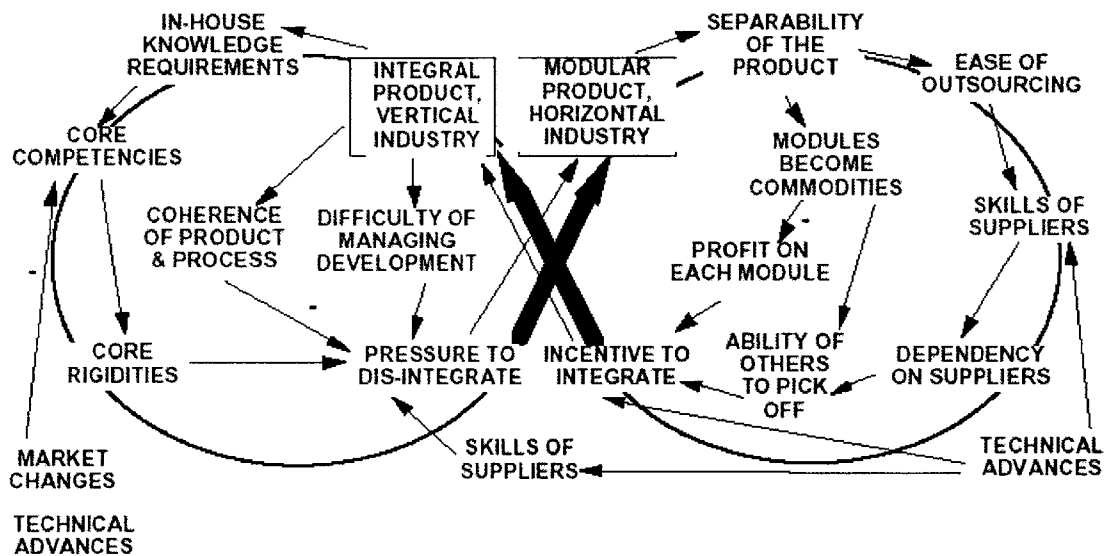


Figure 4: The Fine / Whitney "Double Helix"²⁰

Fine’s follow-up work on the double helix further enumerates the forces of integration and disintegration, summarized in Table 4 below:²¹

¹⁹ The shape more closely resembles a figure-8, however, the “double helix” nomenclature, like many names in the domain where science meets business, has stuck despite the obvious misuse. It seems to have stemmed from an attempt to relate the study of business and biology.

²⁰ From [Fine and Whitney, 1996]

²¹ [Fine, 1998]

Incentive to Integrate	Incentive to Dis-integrate
Technical advances in one subsystem can make that the scarce commodity in the chain, giving market power to its owner	The relentless entry of niche competitors hoping to pick off discrete industry segments
Market power in one subsystem encourages bundling with other subsystems to increase control and add more value	The challenge of keeping ahead of the competition across the many dimensions of technology and markets required by an integral system
Market power in one subsystem encourages engineering integration with other subsystems to develop proprietary integral solutions	The bureaucratic and organizational rigidities that often settle upon large, established companies

Table 4: Forces in the Double Helix

The next three chapters will discuss power management as a technical problem that cannot be solved solely within a module as a new driving force of industry integration.

These would best fit in the double helix model as a new integration incentive force: “the inadequacy of a modular architecture to overcome technical barriers.”

2.4 Moore's Law: Cultural Icon, Industry Driver, and Corporate Monument

2.4.1 The Original Law

One of the most important influences in the microelectronics and computer industries over the past 40 years has been Moore's Law. This idea that the circuit density and, hence, speed of microelectronics will continue to double every 18 months is based loosely on a statement made by Intel founder Gordon Moore in a 1965 article. Part observation, part prognostication, his exact quote is, "The complexity for minimum component costs has increased at a rate of roughly a factor of two per year. Certainly over the short term this rate can be expected to continue, if not to increase. Over the longer term, the rate of increase is a bit more uncertain, although there is no reason to believe it will not remain nearly constant for at least 10 years."²²

In the 1975 IEEE International Electron Devices meeting, Moore revised his statement to the number of transistors on a chip doubling every two years. Somewhere along the way, it became 18 months, as this is the approximate actual time period of density doubling over the past 30 years (see Figure 5).

²² [Moore, 1965]

2.4.2 The Real Law

Regardless of the preservation (or lack thereof) of his original wording, Moore's Law has had a profound impact on the computer user's psyche, the industry marketplace, the industry's roadmap, and, thus, computer component product development. Exploring each of these aspects to the complex phenomenon that is Moore's Law helps to set the context for the complex industry dynamics that are currently taking place in the computer industry.

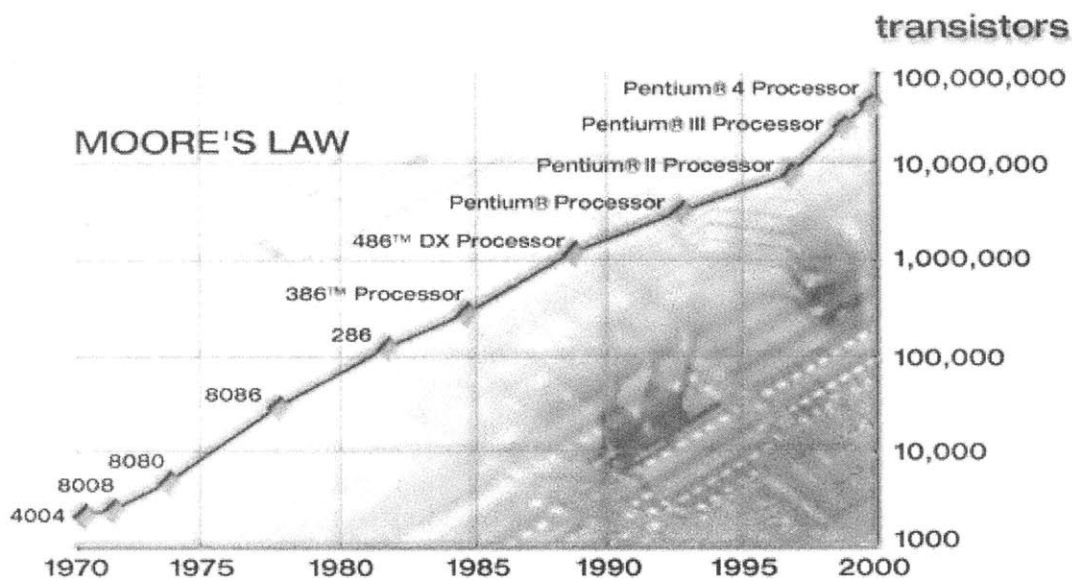


Figure 5: Circuit Density of Intel Microprocessors 1970-2000²³

Although quite obviously not a “law” in the traditional scientific sense, Moore's Law is at once a cultural icon, industry driver, and corporate monument. Even though it is often

²³ Figure from: <http://www.intel.com/research/silicon/mooreslaw.htm>. Accessed December 2003.

incorrectly quoted to mean the doubling of processor speeds every 18 months, this is the standard which consumers have come to expect from the marketplace. See Figure 6 for a plot of both raw processor speed (external clock rate in MHz) and “true” processor speed (amount of processing the chip can do in one second in Millions of Instructions Per Second or MIPS) over time. A regression analysis of both of these lines yield a slightly different exponential trend than the 1.5 years that is most frequently quoted: that raw speed (in MHz) over time doubles only every 2.9 years ($R^2 = 0.95$) and that the true speed (in MIPS) doubles only every 2.1 years ($R^2 = 0.95$).

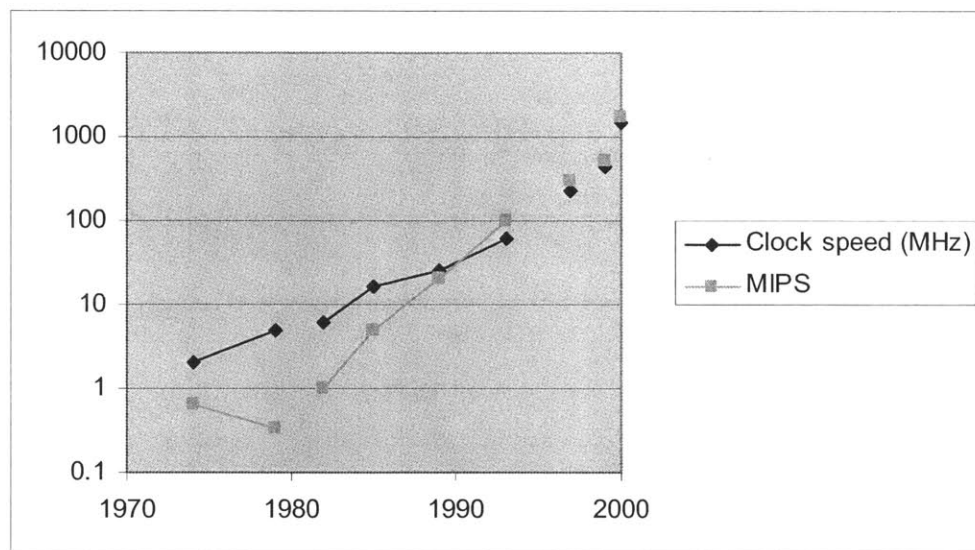


Figure 6: Speed of Intel Microprocessors 1970-2000²⁴

²⁴ Data obtained from: <http://computer.howstuffworks.com/microprocessor1.htm>. Accessed November 2003.

2.4.3 The Self-Fulfilling Prophecy

Furthermore, Moore's Law (combined with the marketing strength of "Intel Inside") has made the processor type and speed the most important parameter in consumers' personal computer purchase decisions. The average consumer virtually ignores important parameters such as memory size and frontside bus speed, which can affect the computer system performance just as much as the processor speed.

The aforementioned market influences are important because these, combined with industry roadmaps and software developers, help to make Moore's Law a self-fulfilling prophecy. The International Technology Roadmap for Semiconductors is a document prepared by a consortium of chipmakers, equipment/material suppliers, and research institutions that lays out the next 15 years of semiconductor technology in order to help guide investments and research efforts "as [they] strive to extend the historical advancement of semiconductor technology and the integrated circuit market."²⁵

The document even goes so far as to mention Moore's Law by name. Thus, the law's influence has become so great that it actually has become the worldwide target for technology development.

Another positive feedback loop that helps to guarantee the persistence of Moore's Law has been the software industry. Wirth's Law, which states: "Software gets slower faster

²⁵ *International Technology Roadmap for Semiconductors*, 2003 Edition, Executive Summary, page i.

than hardware gets faster,"²⁶ sums up why Intel has been able to push MHz onto the marketplace. In order to run the more recent, more bloated software applications with even more unneeded features than the last version, a faster processor is deemed necessary. However, in reality, only some additional memory is probably sufficient for most users who will never tap into "bloatware" such as the visual basic features of Microsoft Word™.

2.4.4 The Method Behind the Madness

But why all the emphasis and hype surrounding Moore's Law? With an 82.9% market share in x86 (IBM-compatible) microprocessors in 2003, Intel's strategy is aimed just as much at expanding the overall PC market as it is at capturing market share from its chief rival, AMD (15.4% market share).²⁷ Therefore, Intel attempts to drive a 3-year replacement cycle for desktop PCs and a 2-year replacement cycle for notebook PCs.

Much of the Intel marketing machine over the past 10 years has been focused on pushing speed upgrades onto consumers who already own an older PC at the rate of Moore's Law. Convincing consumers that these upgrades are necessary becomes even more critical for

²⁶ Nicklaus Wirth is a retired professor from the Swiss Federal Institute of Technology in Zurich. Also commonly stated as, "Intel giveth and Microsoft taketh away."

²⁷ Dean McCarron, principal analyst at Mercury Research as quoted in "Intel, AMD comfy during fourth quarter", on CNET News.com, February 3, 2004: <http://news.com.com/2100-1006-5152255.html>. Accessed February 2004.

Intel to maintain their profits in the face of decreasing average selling prices for state-of-the-art processors.

This strategy has served Intel very well as they have obtained their dominant position in the industry. However, the concept of trying to “push” faster and more powerful microprocessors onto the customers is very much in contrast with a fundamental tenet of the Lean Enterprise: the “pull” philosophy.²⁸ This makes Moore’s Law, in the language of lean, a “corporate monument”—an ideal or policy that has been passed down and seems too powerful to change. Perhaps this is summarized best by Michael Malone’s February 2003 article in *Red Herring*:

Industry watcher Donald Luskin noted earlier this year that even Intel is finding itself being slowly crushed by Moore's law. He pointed out that just to keep its revenue level, Intel must convince its customers to double their power every 18 months or to stick with its current offerings and find twice as many customers.

That was a lot simpler five years ago, when the economy was strong, much of the market was still untapped, and wafer fabs, which double in price every four years (jokingly called "Moore's second law"), were a lot cheaper.

"That's why Intel's revenue growth just imploded, even as they ship record volume," said Mr. Luskin. "In this deep recession, Intel just can't keep up with the law named after its founder." Mr. Luskin wasn't talking about ever-more-powerful Pentiums—Intel can do that—but ever-hungrier customers. Even Intel can't manufacture them.

As Mr. Schmidt points out in his notes, with Intel's research and development costs doubling every 18 months (apparently R&D follows Moore's law as well), in another 20 years the company's R&D costs will be \$31 trillion annually. Something must give long, long before then.

But give the last word to Mr. Moore himself, who once said, "Obviously, you can't just keep doubling every couple years. After a while the numbers just

²⁸ [Murman, *et al*, 2002]

become absurd. You'd have the semiconductor industry alone bigger than the entire GDP of the world."

Clearly, this "manufacturing" of customers is not indefinitely sustainable, but even if it were, the industry is about to hit some major fundamental physical challenges in the form of power management for mobile computers. The next chapter will set the stage with some background information on the history of the mobile computer industry, while the subsequent one will detail the technical barriers to maintaining Moore's law facing the industry today.

3 The Mobile Computer Industry

Since mobile computers (some would say there are still no truly mobile computers) have not been around for long, it is useful to first briefly look at the history of the computer industry structure before mobile computers. Then, the state of the industry when mobile computers entered the picture will be explored. This is essentially where the industry is today, but, as this thesis proposes, this has been slowly changing and is expected to continue to do so going forward.

3.1 A Brief History of the Computer Industry's Structure

Most accounts of the history of the computer industry tell the story of the architecture changes from the standpoint of the type of system delivered: the design of room-sized calculation machines made of vacuum tubes in the 1940s; the emergence of IBM and its mainframe computers in the 1950s and 1960s; the arrival of the minicomputer in the 1970s; the personal computer in the 1980s; the internet in the 1990s²⁹; and now the era of mobile/ubiquitous computing that has yet to be completely defined. However, important to this discussion is the change in the underlying modularity of the systems delivered, and this story has two major turning points.

²⁹ [Baldwin and Clark, 2000]

First, in 1964, IBM introduced the System/360, the first truly modular design. Before this time, designs were not modular; a business bought a computer by choosing among a group of integral mainframe systems with no mix-and-match capability of components. However, the System/360 was a completely proprietary modular design with all components initially designed and manufactured by IBM. This was strategically important to IBM because, “once a user was committed to the System/360 family of machines, a complex instruction set and even more complex operating system made it expensive to switch to another vendor’s system. This was not a big problem for most users, however, because the breadth of the family, made possible by its modular design, made such switches unnecessary.”³⁰ The proprietary nature of the System/360 meant that although the design was modular, the industry structure was still vertical. Figure 7 illustrates the situation, which lasted approximately until the mid-80s. However, the seeds were sown for the growing number of component manufacturers to “capture the rents” when the second major turning point occurred.

³⁰ [Baldwin and Clark, 2000]

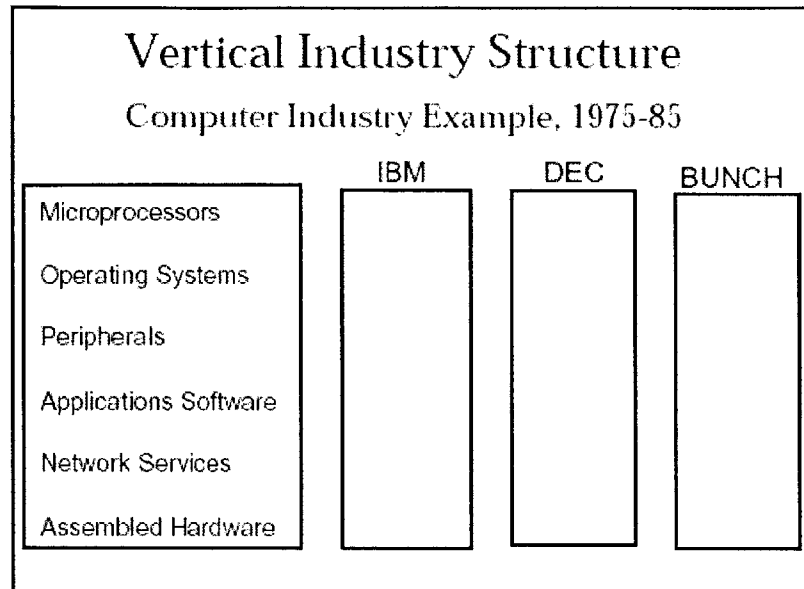


Figure 7: Computer Industry Structure Prior to 1985^{31, 32}

In August of 1981 and in response to the growing market for personal computers fueled by successes such as the Apple II³³, IBM released the PC. It was not the first personal computer, but it was the first IBM computer with an open architecture.³⁴ This means that the components were made by third-party companies using freely available interface standards. This is significant because, for the first time, there could be open competition

³¹ From [Fine and Whitney, 1996] The original is attributed to Andrew Grove, CEO of Intel from 1987-1998.

³² DEC is Digital Equipment Corporation, BUNCH is an acronym for the five second-tier computer suppliers: Burroughs, Univac, NCR, Control Data, and Honeywell

³³ [Fine, 1998]

³⁴ "Inventors of the Modern Computer: The History of the IBM PC - International Business Machines", About.com: <http://inventors.about.com/library/weekly/aa031599.htm>, Accessed Feb 2004.

among the component manufacturers. This led to higher quality components from companies such as Intel and Microsoft and loss of control in defining the value-add components by IBM. In fact, the market capitalization of Microsoft passed IBM's for the first time in 1995.³⁵ The horizontal industry structure created is shown in Figure 8. It had its roots in the modular system architecture of the System/360 and completed the trip around to the horizontal / modular side of the double helix with the advent of the open architecture PC, which "let Intel inside."

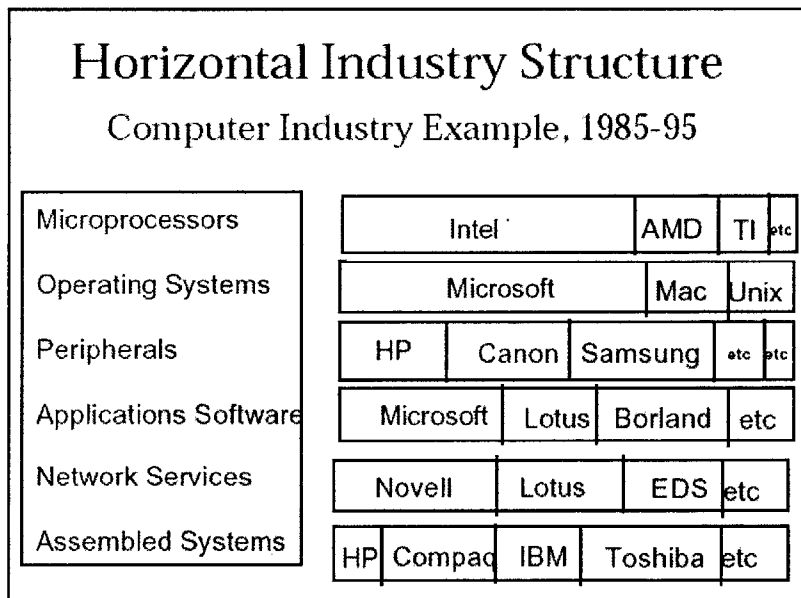


Figure 8: Computer Industry Structure After 1985³⁶

³⁵ [Baldwin and Clark, 2000]

³⁶ From [Fine and Whitney, 1996]

3.2 Along Comes the Mobile

From niche market beginnings in the Grid Compass, a computer used by NASA on the space shuttle program in the early 1980's that was one fifth the weight of any model equivalent in performance (see Figure 9), the mobile computer industry has exploded into a nearly \$100 billion per year industry. In fact, laptop sales eclipsed U.S. retail dollar sales of PCs for the first time in May 2003.³⁷



Figure 9: The 1979 Grid Compass (left) and the 1989 NEC UltraLite (right)

In 1989, the NEC UltraLite (also pictured in Figure 9) was released, considered by most to be the first "notebook style" computer, resembling most of the machines sold today.³⁸

The trend apparent in the form factor of these mobile computers is that the profile and hence, internal volume of the machine, becomes less and less. Meanwhile, the power

³⁷ [Alexander, 2003]

³⁸ "History of Laptop Computers ", About.com: <http://inventors.about.com/library/inventors/bllaptop.htm>. Accessed Feb 2004.

needed by the processor and other components has increased at an exponential rate. As is discussed in the following chapter, this limits the amount space for a cooling solution and the volumetric airflow through the cooling device. Furthermore, the desired trend is for lighter machines that run longer without plugging in to an AC power outlet, meaning there is less space and weight available for a portable power source. Both of these design constraints are power management issues, one that deals with heat dissipation and the other with battery life.

In the early days of personal computing, under the DOS and CP/M operating systems, there was no power management—computers either used 100 percent of their power requirements or were switched off. Personal computer power management history dates back at least to 1989, when Intel shipped processors with technology to allow the CPU to slow down, suspend, or shut down part or all of the system platform, or even the CPU itself, to preserve and extend battery life. Thus, the mobile computer platform started the drive for PC power management. Early attempts at mobile PCs, including the Grid Compass and other “luggables” with heavy cathode ray tubes, drew so much power that they had no alternative but to plug into external AC power sources. In the late 1980s, the development of low cost, reliable liquid crystal displays (LCDs) made battery powered laptops possible, and once hardware technology crossed that line, a sequence of hardware

and software improvements began that have combined to increase the performance and battery lifetime of laptops.³⁹

The first attempt at using component manufacturer cooperation to reducing PC power consumption was the Advanced Power Management (APM) specification, introduced in 1992 by Intel, IBM, and Microsoft. APM was aimed at coordinating the power management activities of the operating system and the BIOS (Basic Input / Output System). However, APM was flawed in a number of ways and gave way to the Advanced Configuration and Power Interface (ACPI) specification of 1997, allowing centralized control of power management by the operating system. “ACPI required Intel and other chipset developers to provide management capabilities in the hardware, Microsoft to implement functionality in Windows, motherboard designers to use the ACPI chipsets and provide the related support, power supply suppliers to implement dual mode supplies, and driver writers to support power management functions.”⁴⁰ Hence, the integrality of design required to deal with power management in mobile computers had already started to increase in the early 90s as is evidenced by the greater amount of cooperation needed among the component manufacturers (i.e. a greater number and complexity of interconnections among modules). In other words, as the problems with high power side-effects become greater, so does the need for integral solutions.

³⁹ [Kolinski, Chary, Henroid, and Press, 2001]

⁴⁰ [Kolinski, Chary, Henroid, and Press, 2001]

3.3 A Short Reprise Before the Punch Line

Now that this thesis has discussed:

- the theory of modularity;
- fundamental differences between information processing and power transfer systems that enable a greater extent of modularity and design automation for information systems;
- the dynamic nature of the integrality and modularity of a product architecture and its effects on the industry structure;
- Moore's Law and its implications for industry and society;
- firm dynamics of the computer industry and its present modular architecture / horizontal structure; and
- new design challenges facing the computer industry with the advent of the mobile platform;

the next chapter will explore the two major forces acting on the product architecture of mobile computers that are paving the way to change the industry structure back to integral / vertical.

4 Power as a Chief Limiter in Mobile Computer Modularity

The Japanese are a tough lot to please when it comes to gadgets. So when NEC Corp.'s stand at the WPC Expo computer trade show in mid-September started to draw crowds like Tokyo's Shinjuku Station at rush hour, it was clear something extraordinary was on display.

It turned out to be a fuel-cell-powered laptop. The 4 1/2-lb. computer can go five hours before its cell needs to be refilled with methanol. That performance beats all but the hardest of laptops running on regular batteries.⁴¹

The above speaks volumes about the two major forces acting on the mobile computer industry structure:

- *Push*: There is a major power crisis looming for mobile computers and in order to continue with their strategy of driving a 2-year replacement cycle for laptop processors, Intel must overcome this system limiter. So, in order to continue pushing MHz onto the market, Intel must find new ways to overcome the power constraints; many of these solutions involve altering other parts of the system than the processor itself and, therefore, require an increased level of product architecture integrality.
- *Pull*: Users are willing to put a flammable liquid into their laptops to get a mere 2 extra hours of operation: this demonstrates that users are starting to value other product features than MHz and are willing to spend their money on ancillary system features, such as battery life. Thus, Intel is forced to design and sell other high-margin components as users become unwilling to spend their money on the

⁴¹ [Kunii, 2003]

cutting-edge microprocessors. By requiring that Intel diversify its method of value delivery from MHz to true mobility (light weight, long battery life, wireless connectivity), the mobile computer users are pulling an increased level of diversification and, hence, vertical integration out of the firm.

As will be seen below, both of these forces will cause a more integral industry structure because the solutions to these problems require a more integral approach to the design of mobile computers. Furthermore, both of these forces are a direct result of difficulties with power management at the microprocessor and the system level. That is, mobile computers are turning into CEMs that transmit significant power. However, this case is different than traditional CEMs in that the significant power is a *side effect* of the information processing function of the mobile computers. Thus, according to Whitney's theory, the mobile computers are at risk of becoming non-modular because of transmitting significant power, but are made even more so because important interactions are occurring across "sneak paths" rather than defined interfaces. The challenge of the mobile computer system designer, then, is to make these significant power interfaces more well-defined so that power can be minimized and released from the densely packed system in an orderly way.

4.1 Pushing Moore's Law: CMOS Scaling and Power

As described in Section 2.4, Moore's Law has fulfilled its promise of cheaper and faster computers over the last 40 years. Engineers and scientists have achieved this primarily through shrinking the fundamental building block of the microelectronic circuit, the CMOS (complementary metal-oxide-silicon) transistor. Shrinking the size alone enables a transistor that switches faster and uses less power per switch at the cost of the new equipment to make the smaller circuits. However, these advantages are meaningless without increasing the number of switches per unit time, or the frequency, that the device performs. Furthermore, the number of devices per unit area is increased, as each transistor now occupies less total area. This allows the increase in "complexity for minimum component costs" that Moore wrote about; more transistors can be cost-effectively put on the same chip.

The above is known as transistor scaling and there are two other key historical scaling factors that are important to the power discussion:

1. V_{dd} , or the drive voltage, is decreased to offset the power added to the chip⁴² by increasing the frequency and the number of transistors. This is because the

⁴² It is also necessary to scale the drive voltage to increase the reliability (continued operation over time) of the transistors. This aspect of scaling started to become prevalent only in the early 90's; drive voltages had been kept relatively constant prior to that time.

switching, or active, power is proportional to the square of the drive voltage. This relationship is given by the equation:

$$P_{ACTIVE} = C_{TOTAL} V_{dd}^2 f$$

where C_{TOTAL} is the total capacitance of the loads (the gates and wires) on the chip being charged and discharged at each switching and f is the frequency of the switching.

2. V_t , or the threshold voltage of each transistor in the given technology, must also be scaled along with V_{dd} in order to maintain the same transistor switching speed. This is because the switching speed is a direct result of I_{DS} , or the drain to source current, of the transistor. The drain to source current when the device is on is proportional to the square of the difference between V_{dd} and V_t , given by this equation:

$$I_{DS} = a(V_{dd} - V_t)^2$$

where a is a constant for any given transistor that depends on the physical properties of the silicon and the dimensions of the transistor gate and oxide thickness.⁴³

However, decreasing V_t has another side effect that has not been important until recently. This increases the non-switching, or standby, power that the transistor consumes while it

⁴³ [Gray and Meyer, 1993]

is not performing any useful function. The predictive equations for standby power are quite complex⁴⁴, but this quantity can be measured empirically by:

$$P_{STANDBY} = I_{sb}V_{dd}$$

where I_{sb} is the standby current produced when the chip is powered on but not switching.

Figure 10 illustrates the relationships described above in a causal loop diagram.

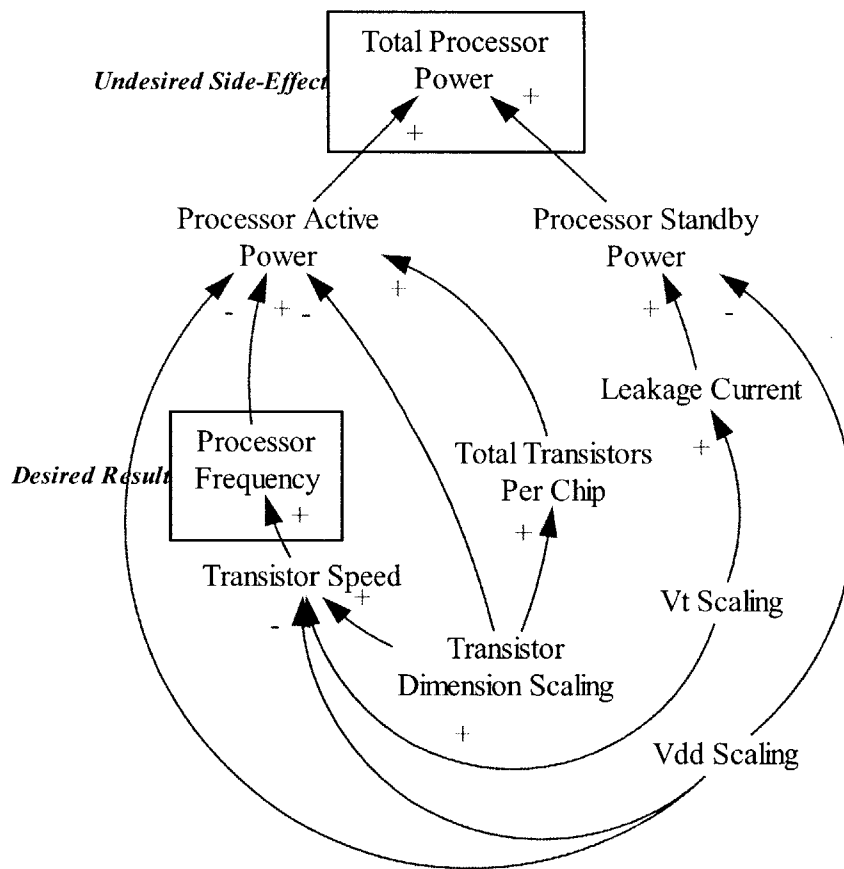


Figure 10: Causal Loop Diagram of Traditional Scaling's Effect on Power

⁴⁴ For a thorough discussion of transistor leakage (standby current) mechanisms, see: Roy, Mukhopadhyay, and Mahmoodi-Meimand, "Leakage Current Mechanisms and Leakage Reduction Techniques in Deep-Submicrometer CMOS Circuits", *Proceedings of the IEEE*, Vol. 91, No. 2, February 2003.

Thus, there is a tradeoff that must take place among these scaling parameter as each new successive technology generation is architected. Table 5 below describes the typical scaling factors that have been used.

Parameter	Scaling Factor	Effect on Power
Transistor Dimension	0.70	decrease
Frequency	1.43	increase
Number of Transistors	2.00	increase
Drive Voltage, V_{dd}	0.75	decrease
Threshold Voltage, V_t	0.75	increase

Table 5: Recent Approximate CMOS Scaling Factors

The standby power had been orders of magnitude less than the active power until recently. Figure 11 illustrates the microprocessor power trends that historical CMOS scaling has produced.

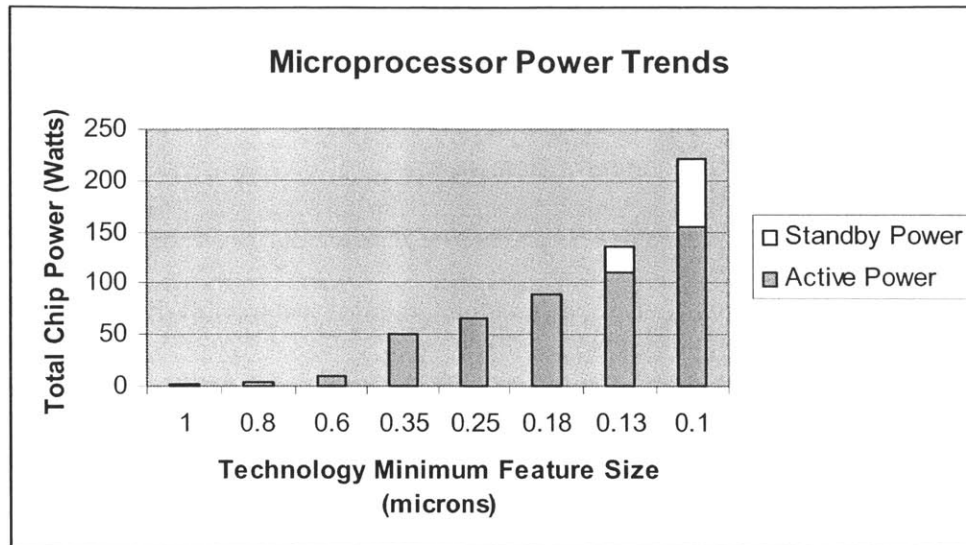


Figure 11: Emergence of Standby Power as Significant Design Parameter^{45,46}

The exponential trend in Figure 11 illustrates that total power and standby power are becoming significant design parameters for microprocessor. In the words of Kerry Bernstein of the IBM Watson Research Center in his keynote address at the 36th International Symposium on Microarchitecture in December of 2003, “Power has become the predominant limit to scaling; new technologies have only limited ability to mitigate power.”⁴⁷ The next two sections show this to be the case for the two main power management issues for mobile computers: cooling and battery life.

⁴⁵ [Rusu, 2001]

⁴⁶ [Thompson, Packan, and Bohr, 1998]

⁴⁷ [Bernstein, 2003]

4.2 Peak Power Density and Cooling Costs

The real issue is not that the power itself is increasing at an exponential rate, but that power dissipation produces heat, and the heat adversely affects the performance of the chip. Each transistor and wire on the chip acts as a resistor that produces heat when a current passes through it. The amount of heat generated is proportional to the resistance of the circuit element, the amount of current going through the element, and the length of time that the current is going through it. Specifically, the heat produced, measured in Joules, is given by the equation:

$$\text{Heat} = I^2 R t$$

where I is the current in Amperes, R is the Resistance in Ohms, and t is the time in seconds.

Heat adversely affects the performance of transistors and, hence, the microprocessor in a number of ways and the effects are both long-term and short-term. Over time, high device temperatures cause hot carrier degradation (HCD), which shifts the V_t and lowers the I_{DS} of the transistors, resulting in slower performance and higher standby currents. HCD can even eventually lead to total device failure.⁴⁸ To make matters worse, HCD increases as device dimensions decrease with transistor scaling. Another long-term effect of heat is electromigration, which is the movement of metal in wires that sustain high

⁴⁸ [Mahapatra, Parikh, Rao, Viswanathan, and Vasi, 2000]

currents and temperatures. This can eventually lead to open connections (disconnected wires) and the failure of circuits.

Second, and most important for this discussion, is short-term performance degradation of the computer system caused by heat. This happens when the processor experiences a throttling or an F_{\max} failure event.⁴⁹ An on-board sensor monitors the chip temperature and halts the execution of instructions if the maximum operating temperature is reached. This is known as throttling and it degrades the operating performance, but prevents the chip from overheating and potentially destroying itself. An F_{\max} failure event occurs when the processor becomes too hot to function normally at the frequency that it is set to run. Transistors and interconnects slow down as temperature increases and this can result in information output by one circuit not being ready in time for use at the input of the next circuit. Efforts are made before the processors are shipped to insure that each chip will rarely experience an F_{\max} failure, but they still do happen, especially since users do not always operate the processors within the rated temperature and frequency range. In other words, a microprocessor designed to operate at 3GHz can only do so if the heat it generates is removed from the system. A summary of the design tradeoffs for removing heat from a mobile computer is illustrated in Figure 12. As can be seen, major limitations are imposed by the cooling solution costs and the small form factor of a laptop.

⁴⁹ [Evans, 2003]

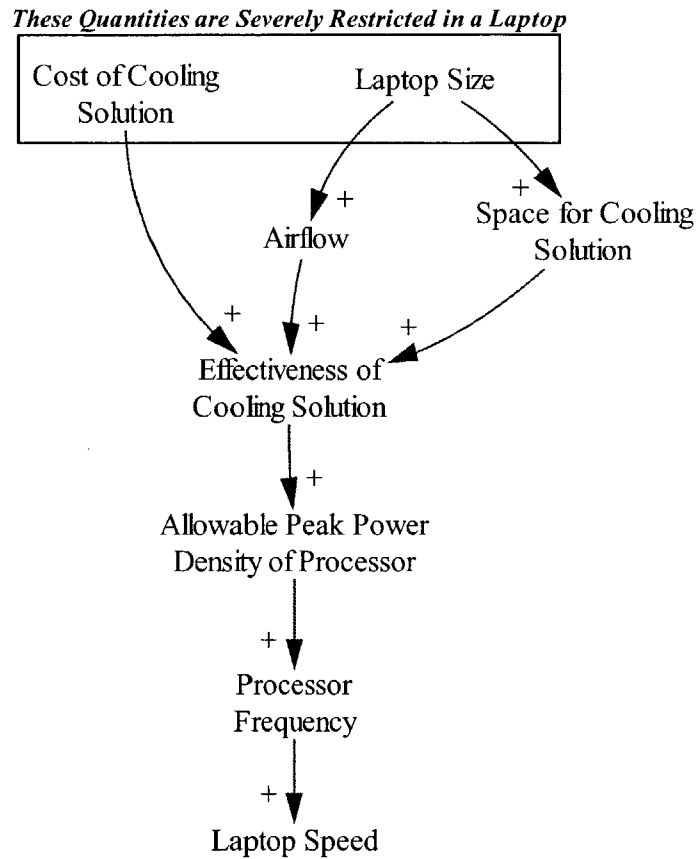


Figure 12: Causal Loop Diagram of Cooling Cost and Size Constraints for a Laptop

Temperature is the result of power density sustained for approximately five or more seconds.⁵⁰ Thus, it is actually the peak power density (heat flux) that is important in determining the type of cooling solution needed for a particular processor. Figure 13 below shows the trend in peak power density of processors and the limits of each of the respective cooling solutions, as well as the cost of each of those solutions.

⁵⁰ [Evans, 2003]

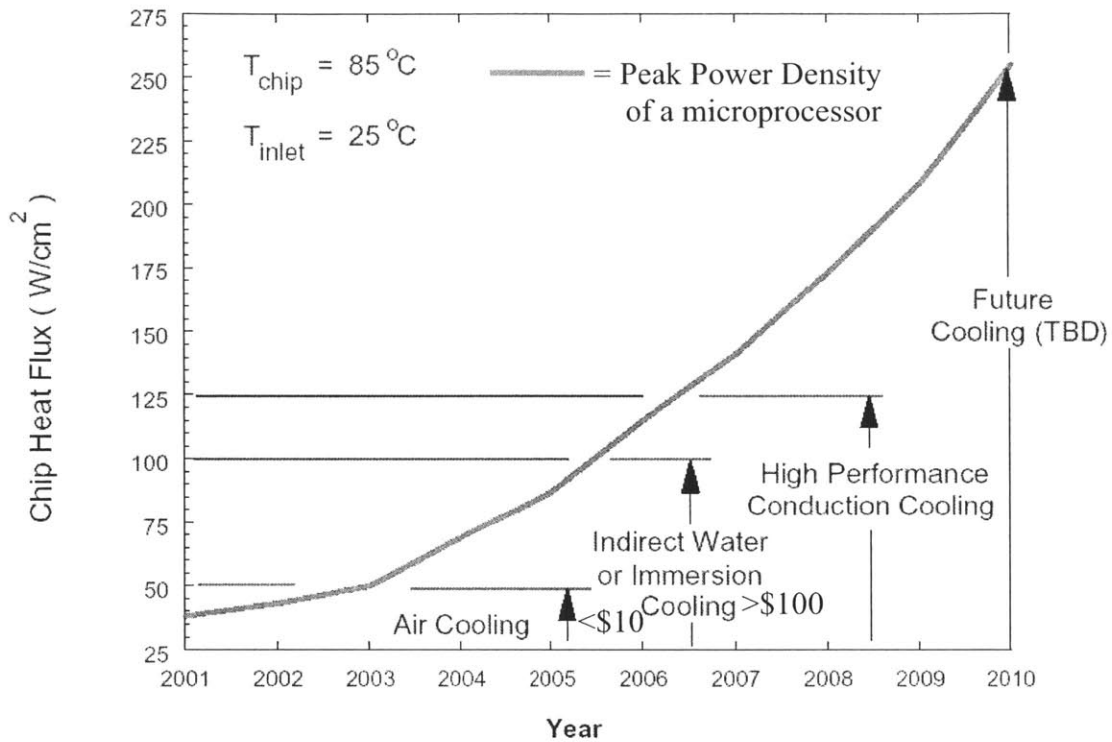


Figure 13: Processor Power Limits for Present Cooling Technology⁵¹

As can be seen, the cost of cooling will become prohibitively large in the near future (within 3 years) given the current peak power density trend. Thus, continuing along the trajectory of Moore's Law is already becoming difficult in the small form factor platform of mobile computers as cost-effective cooling limits have almost been reached. In the words of Intel fellow Shekhar Borkar, "Today you are limited by power. The practical limit is around 75 watts, because if you look at the cost of cooling, that too will start

⁵¹ [Chu, 2003]

increasing exponentially. No one is going to buy a thousand-dollar refrigerator to cool a thousand-dollar PC!”⁵²

The situation has reached a point that the new G5 chip from IBM cannot be put into a Powerbook (Apple’s mobile computer product) because it cannot be cost-effectively cooled. In fact, nine fans are required for the G5 in a desktop platform.⁵³ Clearly, modifications to the Powerbook architecture are needed beyond advanced cooling solutions to solve the problem of putting the latest processor inside.

⁵² interview with Intel Fellow Shekhar Borkar “Delaying 'Forever' — The Future of Moore's Law.” <http://www.intel.com/labs/features/mi03031.htm>. Accessed January 2004.

⁵³ [Salkever, 2003]

4.3 Market Pull: Average Power and Battery Life

Equally important to the mobile computer user is the separate but related power management issue of battery life. Again there is a tradeoff that occurs, but this one, dealing with supplying portable power to a laptop, has to do with limiting size and weight rather than limiting temperature.

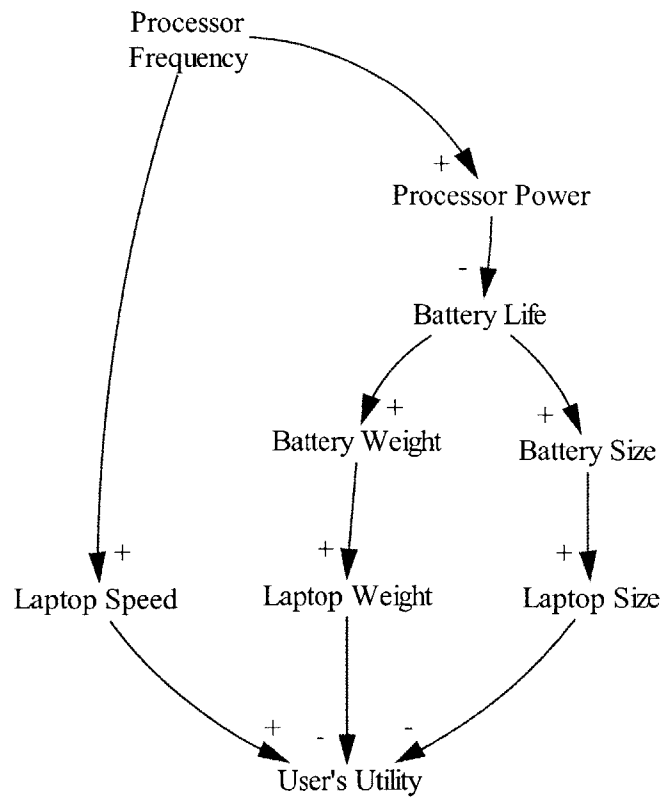


Figure 14: Causal Loop Diagram of Laptop Speed / Weight / Size Tradeoff

Batteries are rated to support an amount of average power for given amount of time. This is measured in Watts x Hours. Thus, the higher the average power of the entire computer

system (including the processor), the shorter “unplugged” operating time the laptop will have. However, more WxH means a battery is heavier and larger. The causal effect of these tradeoffs on the user’s utility is shown in Figure 14.

The current typical lithium-ion battery technology provides 24 to 72 WxH of capacity. The other components in a mobile platform except for the processor consume about 12W of average power.⁵⁴ Thus the 12W can be added to the expected future average power trend of microprocessors to yield a graph of system power over time, as shown in Figure 15. The starting point is the 1W average power of the Pentium M⁵⁵ of today and the trend follows the same exponential presented in Figure 11. The total unplugged operation time can then be computed for both ends of the capacity envelope (24 and 72 WxH), also shown in Figure 15. As can be seen in the figure, the life of even the best performing batteries of today will come nowhere near meeting the power demands of tomorrow’s laptops unless the power trend is halted. Within the next 10 years, even the 72 WxH batteries, the heaviest and most expensive, will last less than 2 hours. This may seem like a long time horizon, but bear in mind that the current solution is already not satisfactory to most laptop users: the current choice is between a lightweight machine with about 2-3 hours of battery life or a heavier machine with up to 6 hours.

⁵⁴ [Gochman, et al, 2003]

⁵⁵ The Pentium is the runaway leader in performance for power of mobile processors available today and it has already begun to sacrifice speed for power, so it is a realistic starting point for a “new” exponential trend that will not have a discontinuity due to a shift in Intel’s strategy of offering a power optimized processor.

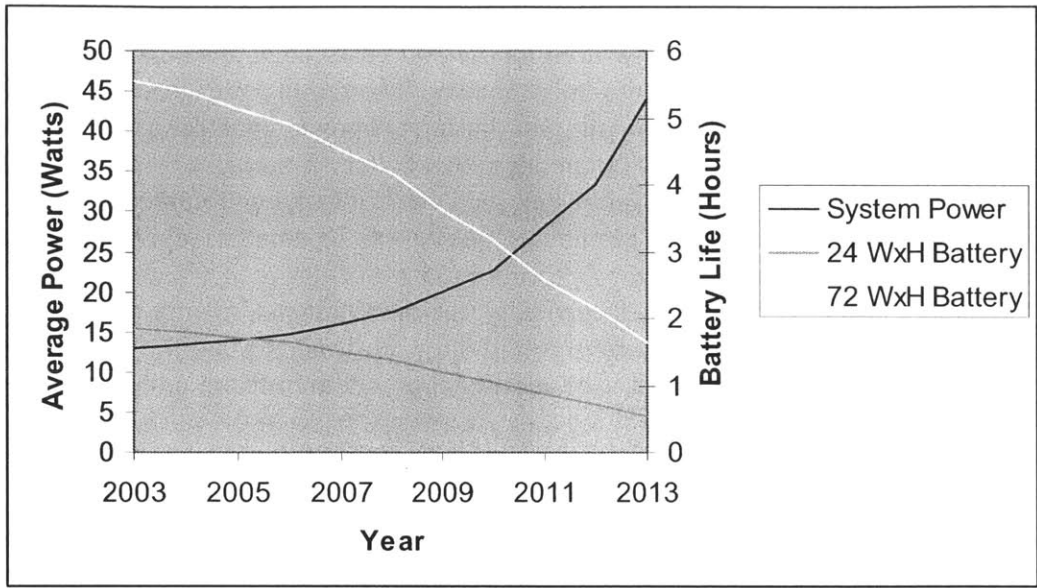


Figure 15: Processor Power Limits for Present Mobile Supply Technology

The impending doom of Figure 15 is spelled out pretty well by this article from Newsweek.com mourning the absence of a new battery technology in the face of an urgent need for one:

Let’s face it: the biggest challenge for advanced road warriors today isn’t finding the nearest WiFi hotspot. It’s making sure that the battery doesn’t run out in the middle of a crucial download. Ironically, in our headlong rush to create sophisticated untethered computing, the most problematic technology turns out also to be the oldest: those nondescript metal cylinders that never seemed to be included with our Christmas toys.

Suddenly, however, the quest for long-lasting portable power is on everyone’s mind, from blue-chip Silicon Valley venture capital firms to Japanese giants like Sony and Toshiba. And it looks like there are only two options: continued improvements in existing technology, or a major breakthrough, most likely miniature fuel cells that produce electricity using methanol as fuel.

Batteries are a very old technology—2,000 years ago, jewelers in Baghdad apparently used simple batteries to electroplate their creations with thin layers of gold or silver. The technology was reinvented early in the eighteenth century, when Alessandro Volta demonstrated the first Voltaic cell for Napoleon

Bonaparte, giving us both the concept of the battery and the name of the unit by which electric potential is measured. In the 200 years since, the fundamental concept hasn't much changed—only the materials within have evolved, growing increasingly exotic and culminating in the current power champion, lithium-ion.

But at present there's not another new battery ingredient on the horizon that can substantially beat lithium-ion. And that's a major problem. Moore's Law states that the complexity of circuits doubles every 18 months, while on average battery capacity increases only 5 percent to 10 percent a year. Generally speaking, the more complex a circuit—specifically, the more transistors on a microprocessor—the more current it will draw. Now add to that the new demands of wireless computing, which uses additional power to transmit and receive, and it's pretty clear we're approaching a major energy crisis in the portable world.⁵⁶

This power management issue sets the stage for the second major force acting to limit modularity of mobile computer design: the pull of the market. This topic was explored at the introduction of this chapter, but is perhaps best summed up industry observer Paul Boutin:

The confusion over the meaning of Moore's Law led some industry watchers to raise their eyebrows at Intel's new, unspoken shift in strategy: With the launch of the Centrino mobile chip set, Intel has abandoned the shorthand definition of Moore's Law. For the first time in its history, Intel isn't touting the clock speed of a new CPU. The Pentium M central processor at the core of Centrino ticks over at a lazy 1.6 gigahertz, 20 percent slower than last year's mobile version of the Pentium 4. But despite its slower clock speed, Centrino doesn't mean that Intel has given up on Gordon.

With Centrino, Intel proves that all those transistors can be used for lots of things, not just sheer speed. Pentium M's all-new design beats the P4's count of 54 million transistors on one chip with a new high of 77 million. It's not double the old count, but it's a big leap. Instead of cranking up the clock speed and then hunting for reasons for PC owners to upgrade, Intel has turned around to meet its customers' biggest grievance: laptops that run out of juice. The extra transistors on the Pentium M bring more memory cache onto the same chip, saving precious battery power. Other new circuits are dedicated to controlling and conserving

⁵⁶ [Rogers, 2003]

power. Centrino's built-in Wi-Fi is handled not by the Pentium M but on a separate chip. Still, integrating it next to the CPU reduces battery drain.

If Centrino-equipped laptops really run five, six, or eight hours on one battery charge, as claimed, that will be a doubling of another sort. It's unlikely that the doubling of battery life will become the next shorthand meaning of Moore's Law, but for laptop users, it's something they need more than another couple of gigahertz.⁵⁷

In other words, mobile computer customers are starting to value other things than MHz and Intel is trying to deliver these things. Even though this seems to defy the colloquial shorthand of Moore's Law, "doubling the clock speed every 18 months," it really meets the criteria of Moore's original wording that the increased complexity for the same cost will result in performance improvements (if performance improvements are measured in other ways than just clock speed).

The power management issues outlined above are looming on the near-term horizon and are already beginning to have a large impact on the industry's research and development efforts. Technically feasible, cost-effective solutions must be found within the next few years. Table 6 below provides a list of possible options and the modular or integral nature of the solution. Note that here, microarchitectural solutions or "integrating" more logic functions onto a processor, are considered a modular solution since the designers are solving the problems created by the module (the microprocessor) by altering only the "hidden" design parameters contained completely within the module. Note also that

⁵⁷ [Boutin, 2003]

operating the processor at a slower frequency essentially commoditizes high clock speed designs. Thus, as mentioned in the introduction to this chapter, Intel must get into other “value-add” features of the processor and other components of the computer system in order to continue to be profitable. Successfully implementing this new strategy requires an increased level of vertical integration.

Description of Solution	Integral (I) / Modular (M)
Semiconductor process improvements	M
Microarchitectural solutions	M
Operating the processor at a slower frequency	M
More power-efficient software	I
Advanced cooling mechanisms / battery technologies	I
Lower power components (screen, hard drive, etc)	I

Table 6: Options for Mobile Computer Power Management Solutions

For the most part, all of these options are presently being pursued and power management problems are only getting worse as Intel continues to push megahertz (although not as vehemently in the mobile platform as before) and users increasingly desire longer battery life, lighter weight, and smaller size from their laptops. What this means for system and component designers alike is that all options must continue to be pursued if the computer industry hopes to deliver a valuable product. This, in turn, means further integrality of design for mobile computers.

5 A Framework for the Dynamic Modularity of Industries

Now that the case has been made for power management as a technological limiter to modularity of design in mobile computers, this chapter will discuss the business implications. It is therefore useful to develop a framework for the relationship between these dimensions (business and technology) as they relate in the domain of modularity.

5.1 The Framework

The framework below relates modularity of design and modularity of the industry by measuring both of these quantities for a particular industry at a particular point in time along the continuum of each of two axes (see Figure 16). Thus, the more modular the architecture of a product produced by an industry is, the farther to the right on the x-axis it is. Likewise, the more that an industry's firms are organized in a modular fashion, the closer to the top of the chart it will be. Recall the definitions presented above for each of these type of modularity:

product architecture modularity — the degree to which a product can be broken into components that can be individually designed and manufactured according to predetermined design rules

industry structure modularity — the degree to which those individual components are designed and manufactured by different firms

The following two sections describe the terms introduced in Figure 16 in greater detail.

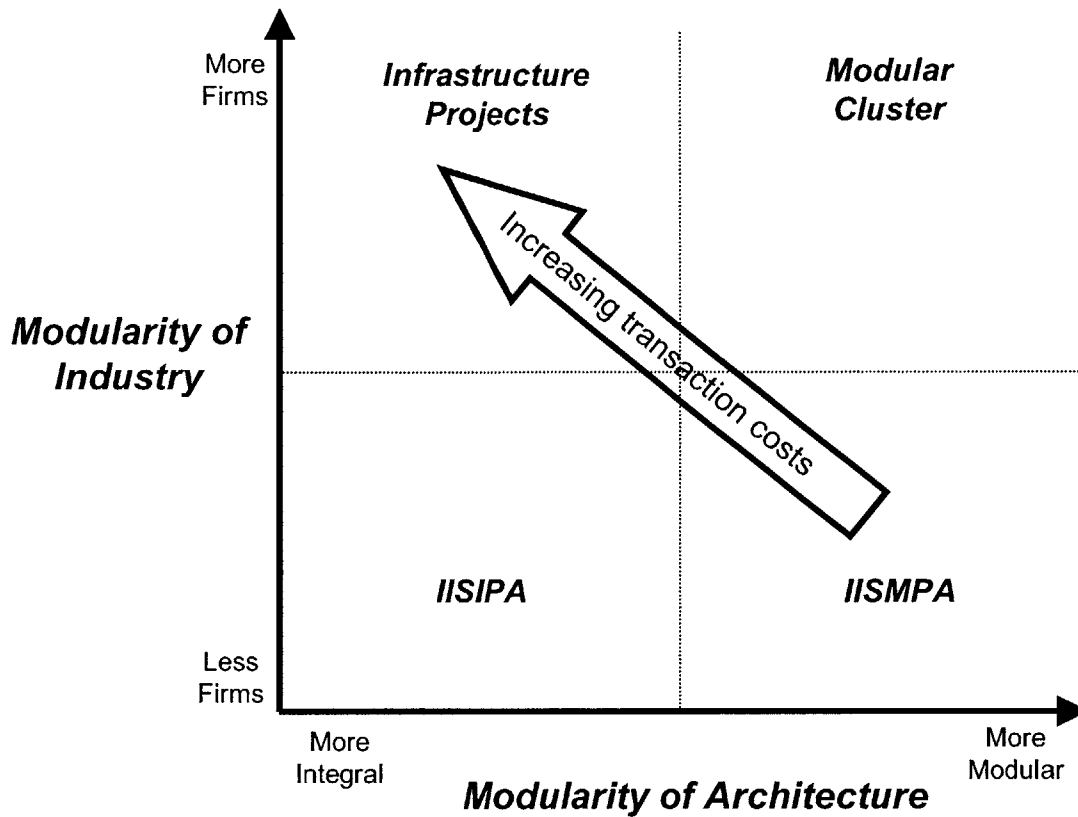


Figure 16: Modularity Framework⁵⁸

(IISIPA = integral industry structure / integral product architecture;
 IISMPA = integral industry structure / modular product architecture)

⁵⁸ The relationship to the double helix model should be apparent and this figure is definitely a cousin of Figure 8.1 in *Clockspeed* [Fine, 1998]. However, it is the combination of the two models and the addition of the transaction costs that allow the novel “free body diagram” approach used here.

5.2 Transaction Costs

Another concept that will be important to the application of this framework is that of transaction costs. Simply stated, transaction costs are the costs incurred in transferring material, resources, or information. As is pictured in Figure 16, these costs increase as the product architecture becomes more integral and the industry structure becomes more modular. The reasons for this are as follows:

- The more integral a product architecture, the more ill-defined the interfaces are between chunks, and the more communication that must occur between designers of the separate chunks. Also, more unexpected system-level effects will happen when the chunks are put together, resulting in more iterations (and opportunity for transaction costs) that must occur to produce a working system.
- The more modular an industry, the more the chunks are designed by separate firms and, thus, the more expensive the communication is that must take place. That is to say, transaction costs across the boundary of firm are higher than if the transaction was to take place completely within the same firm. Transaction costs are important in that they are an “invisible” third force acting on the trajectory of most industries at any given point in time. This point will be elaborated further in applying this framework to the mobile computer industry.

5.3 The Four Quadrants

Also labeled in Figure 16 is the name of each quadrant. The bottom half of the graph contains, for lack of more elegant names, “integral industry structure / integral product architecture” (IISIPA) and “integral industry structure / modular product architecture” (IISMPA). These are fairly self-explanatory in that they are the lower half of the modularity of industry axis. In other words, the firms in these industries make all or many of the chunks that comprise the industry’s product. The difference between the two is then just whether or not the product architecture is integral or modular. It should be noted here that the modularity of architecture axis does not measure number of chunks or complexity of the product since integral or modular systems can have greatly varying degrees of complexity.⁵⁹ However, for a given product, a more modular design is, in general, divided into more chunks. For example, a handbag may be made more modular by dividing up the design and manufacture of the handle, the fastener, the main carrying compartment, etc. (note also that this is less expensive in terms of transaction costs than an integral product architecture for the handbag), but this “modularized” handbag still has many fewer chunks than a jet engine, which is a highly integral product.

⁵⁹ Number and complexity of interconnections (not chunks) is, in fact, a good measure of the modularity of a product. A “quick and dirty” modularity metric from network theory is (number of chunks) / (number of interconnections), also called the connectivity density ratio. The point here is that “modularity of product” is not an easy thing to measure; indeed, entire theses can and have been dedicated the topic. Note that the number of firms *is* used on the modularity of industry axis because it is, to a first order, a good approximation of how the product is divided up for design and manufacture among firms in the industry.

The purest form of a firm in the IISIPA domain is a company that supplies only one type of raw material. This includes firms like US Steel, DeBeers, and many others that are probably not as well-known to the average end user of products fabricated from these raw materials. However, it may also include firms that develop more complex integral systems almost entirely on their own, such as Nike (athletic footwear) or Pratt & Whitney (jet engines). IISMPA best describes the computer industry from the introduction of the IBM System/360 in 1964 until the early 80s, as described above. Another example of an industry in this quadrant is stereo component manufacturers that make products such as receivers and speakers.

The name for the upper right quadrant is taken from Baldwin and Clark's *Design Rules*—the “modular cluster.” In this type of industry / product architecture organization, a group of firms and markets play host to evolution of modular designs. When a product with a modular design is created in an economy with advanced capital markets, subindustries of firms and markets organized around modules may emerge and evolve in parallel with the module designs themselves. This represents the mobile computer industry over the past 20 years through today.

Lastly, the upper left quadrant is labeled “infrastructure projects.” These are the special case of the framework as this region describes industries that take on projects that would not be profitable for any one economic entity without outside funding, usually from the

government. In addition, these projects are typically undertaken for “the betterment of society,” hence the name of infrastructure. For example, the development of the Joint Strike Fighter is a good example of a large, multi-government funded consortium project that spans many different firms but requires extreme coordination since the architecture of a jet airplane is fundamentally integral. In general, many large defense projects fit in this category as do other infrastructure projects, such as the human genome project. The amount of information to be gathered in the human genome project was too large for any one firm to tackle, but the end product required an extreme amount of coordination among firms and, thus, high transaction costs. The solution to this was government funding, justified by the possible good that can come out of having a map of human DNA for future applications of genetic engineering.

As with any framework, some industries do not fit neatly into one of the quadrants and are hybrids of the industries described above. The auto industry, for example, has a product architecture and industry structure that is both integral (engine and drivetrain; frame, chassis and body all integral chunks made by the OEMs) and modular (headlights, radio, seats, etc. made by suppliers) at the same time.

5.4 Application of the Framework to the Computer Industry

Now that the fundamentals of the framework have been described, it will be applied to the current state of the mobile computer industry given the forces acting on it described above. In doing this, the graph can be used similarly to a two-dimensional free body diagram in physics or engineering mechanics. By taking the vector sum of the forces currently acting on the industry, the resultant trajectory can be predicted. This is illustrated in Figure 17 below.

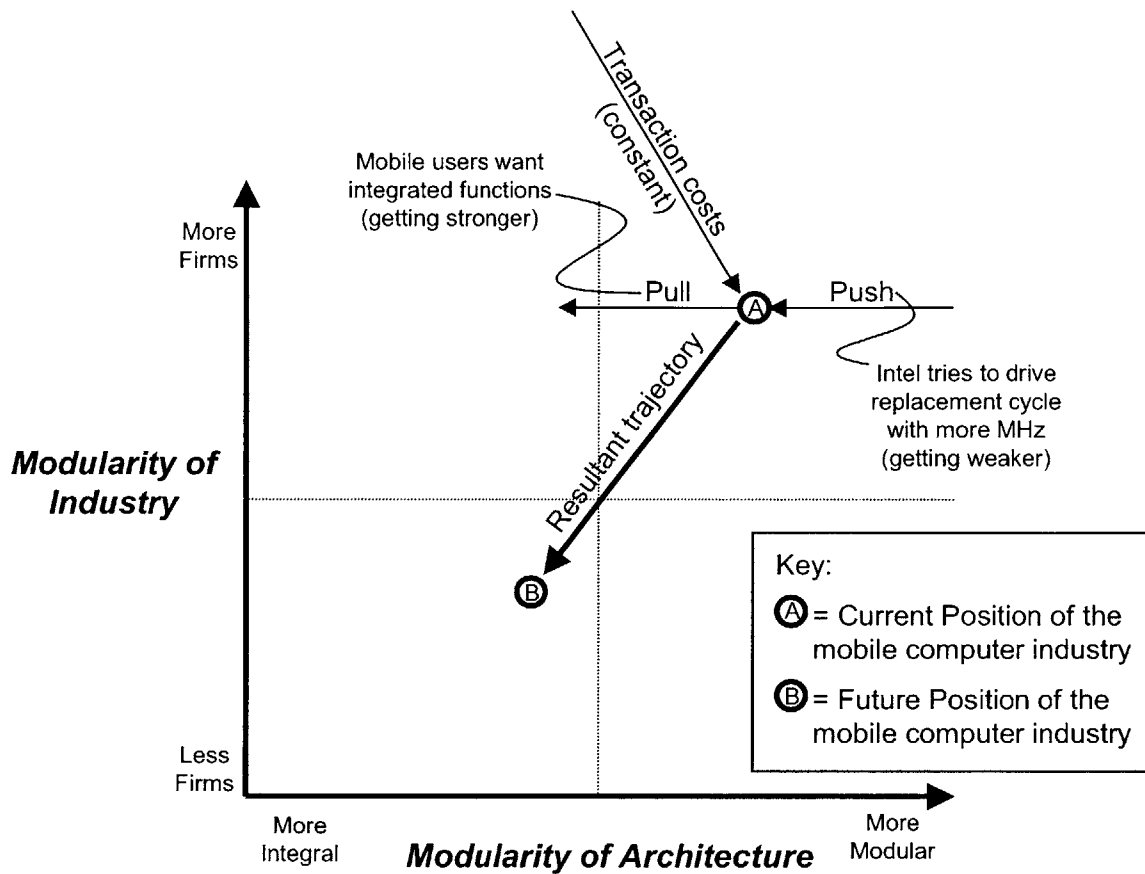


Figure 17: Applied Modularity Framework - "Free Body Diagram"

The mobile computer industry presently sits at point A in the modular cluster quadrant.

There are currently three major forces relevant to this domain acting on the industry's

"point mass." As noted above, there are two forces caused primarily by power

management problems that are causing the product architecture to become more integral:

1. As Intel continues to try to push MHz onto the consumers, breaking down the power density barriers to maintaining the colloquial Moore's Law requires a more integral product architecture.

2. As consumers desire more integrated features and value different product attributes, which include lighter overall weight and smaller size (and, thus, lighter and smaller batteries) and longer battery life, the market is pulling the industry toward a more integral architecture to optimize these valuable design parameters.
3. The third force is the invisible transaction cost force described above. It acts in the direction so as to minimize the transaction costs (down and to the right). This is analogous to drawing gravity on a free body diagram for physics computations.

All three forces are represented by vectors on the graph. Again analogous to physics, the sum of the three component vectors can be drawn to obtain the resultant vector indicating the predicted direction of the industry. Note that the drawing is a simplification in that it is a snapshot of the forces at one particular instant in time. In order to get the true dynamic nature of the industry's resultant trajectory over time, the dynamic nature of the component forces must be captured. However, as noted on the diagram, the transaction cost minimization force remains constant. Additionally, the pull force is increasing while the push force is decreasing. Assuming that these are happening at approximately the same rate, the resultant force over time will remain approximately constant until the industry ends up at point B, somewhere in the IISIPA quadrant.

6 Signs of Verticality

Assuming that this framework is actually modeling reality, signs should begin to emerge that vertical integration is taking place in the horizontal structure of the computer industry. Figure 18 provides a visualization of this trend, illustrating that firms which have become leaders in their respective component will begin to “reach across” the horizontal lines into other modules and eventually become vertically integrated system suppliers. The following examples from Intel, once just a maker of the microprocessor component, demonstrate that this is indeed the case in today’s computer industry.

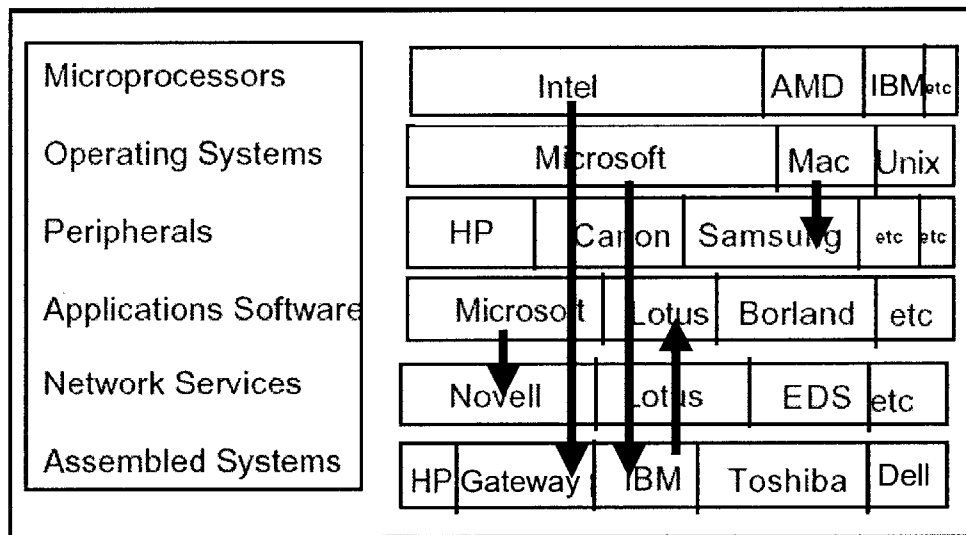


Figure 18: Current Industry Structure (Horizontal Going to Vertical)

One way that Intel has reached across the industry prior to the recent age of power management was in designing entire chipsets. The chipset is the collection of chips that work together on the motherboard to help the processor deliver its functions and

communicate with the rest of the system. Clearly this is not a major parlay into system design, but it was the beginnings of what is today on the cusp of becoming a major vertically integrated force in the industry. Next came the introduction of the Mobile Module in the mid-90's:

Notebook PCs have always represented the greatest design challenges for system makers due to their restrictions on size and weight, and the difficulty in cooling them. To combat this, the trend has been toward more and more miniaturization. Intel is continuing this trend by introducing *mobile module* packaging, which actually incorporates the processor, secondary cache, and chipset into a small module. One could argue that this is almost a motherboard in its own right; it isn't really, but it's pretty close.

Aside from miniaturization, Intel gains from this design tighter control over the interface between the chipset and processor, and also the electrical connections between them, which become more important as performance increases. Of course, it is also a proprietary design, giving Intel marketing advantages.”⁶⁰

The mobile module was clearly a precursor to the amount of vertical integration that is required now that the size, weight, and cooling requirements on laptops have become even more stringent.

Already mentioned in Section 4.3, the introduction of Centrino marks a major step towards more integration, as Intel is beginning to respond to the user's value of ancillary performance metrics, as well as getting into wireless networking chips. As a recent Wall Street Journal article reports, “Centrino also gives Intel more leverage with the computer makers that actually buy its chips. In the past, it offered the manufacturers marketing

⁶⁰ *The PC Guide* (<http://www.PCGuide.com>)

subsidies if they put the 'Intel Inside' logo on their machines. Now, Intel only provides the subsidies if computer makers use the three components of Centrino—the Pentium M, the related chip sets and the Wi-Fi [wireless networking] chips. It marks the first time that Intel had tied marketing subsidies to chips other than microprocessors.”⁶¹ Intel has also developed a new "runs great on Intel Centrino mobile technology" logo to help users identify *software* designed specifically for mobility. Participation in this campaign requires software vendors to “incorporate online and offline capabilities and meet certain power management, performance or connectivity criteria” with their software.⁶²

Additionally, new chipsets with similar integrated power management capabilities are in the pipeline for Intel. The Alviso chipset, slated for introduction in the second half of 2004, offers low power capabilities that will support Intel High Definition Audio, reducing power consumption by allowing the processor to remain in a sleep state while the system plays audio.

Another example of this vertical integration is seen in Intel’s explanation of steps it is taking to solve the power management problem with a “holistic” approach: “Intel researchers, scientists and engineers are taking a holistic approach by looking at every possible place where a variable could influence the power equation. That might be from

⁶¹ [Clark, 2003]

⁶² Intel press release, “Intel Silicon, Platform Innovations Speed Advances In Mobile Computing”, February 18, 2004.

the means, such as the various steps involved in device design and manufacture, to the ends, including the final application, segment or usage model in which the device will be used. This holistic approach is comprised of three basic technology focus areas that are interdependent and build off of each other.” Holistic is just another way to say that they are exploring all options, both modular (the means) and integral (the ends), as can be seen from the large number of system solution listed in Figure 19.

TRANSISTOR		CHIP		SYSTEM		
TRANSISTOR	PROCESS	ARCHITECTURE	CIRCUIT	SOFTWARE	HARDWARE	PACKAGING
<ul style="list-style-type: none"> • High-k gate dielectric • Raised source and drain 	<ul style="list-style-type: none"> • 90 nm process 	<ul style="list-style-type: none"> • Multi-core and clustered micro-architectures • Power optimized micro-architecture • Intel SpeedStep® Technology 	<ul style="list-style-type: none"> • Low-power circuit design techniques • Body bias techniques 	<ul style="list-style-type: none"> • Developer tools: <ul style="list-style-type: none"> - Intel® Vtune™ performance analyzer - Mobility software guidelines 	<ul style="list-style-type: none"> • Voltage regulation technology: <ul style="list-style-type: none"> - Intel® Mobile Voltage Positioning - Intel® Dynamic Voltage Management • Improved power display specs • Thermal design and engineering, including advanced heat sink technology • Reference design solutions 	<ul style="list-style-type: none"> • Techniques to reduce package thickness • Heat spreaders and packaging materials that target "hot spots"

Figure 19: Intel's "Holistic" Approach to Power Management⁶³

Another enabler of further vertical integration is Intel’s participation in the development and support of standards for low power. First APM, mentioned in section 3.2, and now ACPI are power management specifications jointly developed by Intel and other computer component firms. These are significant because it is a first step toward vertical consolidation. As Intel participates more frequently in these initiatives, it begins to

⁶³ Intel website, Research and Development: Designing for Power: <http://www.intel.com/labs/power>.

develop some of the skills of the other component manufacturers (and vice-versa).

Equally important are that the transaction costs of these integral solutions across multiple firms becomes readily apparent, causing firms that do it all themselves to question these costs. Other initiatives relating to integral low power solutions in which Intel is actively participating include these:

- Standards Panel Working Group (SPWG): develops specifications designed to help improve the notebook PC display interchangeability, time to market, and power
- Mobile PC Extended Battery Life (EBL) Working Group: “dedicated to accelerating the achievement of all-day battery life”⁶⁴
- Mobilized Software Initiative: provides a comprehensive set of software, tools, services, architecture specifications and training programs to help design software for mobility, including power management features

Intel also has a large strategic investment division called Intel Capital that began in the early 90’s and now invests in excess of \$1 billion in venture capital for start-up companies that are producing enabling technologies.⁶⁵ Some of the investments in power management technologies by Intel Capital include those listed in Table 7.

⁶⁴ Mobile PC Extended Battery Life Working Group website: <http://www.eblwg.org>.

⁶⁵ Intel website, Intel Capital, <http://www.intel.com/capital>.

Company Name	Company Description
<i>Cap-XX</i>	a developer of supercapacitors, which are a high power, small volume, and light weight power supply that utilizes high surface area carbon to hold charge, instead of the chemical means used in today's battery power sources
<i>Insyde Software</i>	a company based on power management software that extends battery life through ROM- and Windows-based power management products
<i>Iridigm</i>	produces revolutionary reflective displays for mobile applications; that have a very high reflectivity in color, ultra low power, and very low cost
<i>Neah Power Systems</i>	developing silicon-based micro fuel cells
<i>Silicon Wave</i>	provider of low-power, highly integrated RF communication system components for the global Bluetooth wireless market, including single-chip radio processors, stand-alone radio modems, and software solutions

Table 7: Investments in Power Management Technology by Intel Capital

In many cases, these strategic partnerships go far beyond a simple money infusion by Intel, which promises “focused Intel knowledge sharing” that includes access to “Intel’s existing processes, policies, templates, and presentations which portfolio companies can adapt to meet their needs.” These strategic partnerships, then, put Intel in position to more easily acquire the start-ups that are already using their business processes, should they decide they want to become more vertically integrated but are missing key core competencies.

The last class of examples is Intel’s design of enabling platforms and little publicized PC system. These are strong indicators of Intel’s drift, whether intentionally or unintentionally, to the vertical integration domain. Each year at the Intel Developer’s

Forum, Intel introduces reference platform designs to help system designer use Intel's latest technology more effectively. This past year's featured three new mobile computer designs (see Figure 20) as part of the Intel Mobile Internet PC 2004 Platform Vision:

The 12-inch model features EMA⁶⁶ functionality and converts from a laptop to a tablet PC, allowing maximum flexibility to balance office and mobile demands. The 15.4-inch model is designed to enhance worker productivity with fingerprint and smartcard security, built-in array microphones and camera for collaboration, and EMA functionality. The 17-inch Mobile Entertainment PC allows users to communicate and be entertained around the home with a wide-screen display in a sleek, portable design; a wireless Bluetooth keyboard; built-in voice-over-IP handset and remote control; integrated array microphones and camera; and Intel High Definition Audio for high-quality sound.⁶⁷

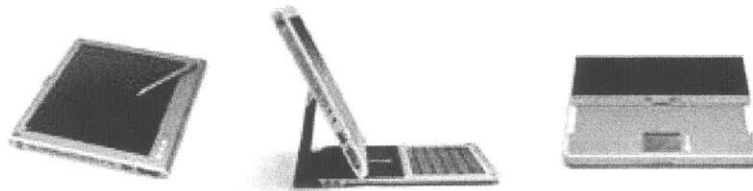


Figure 20: Intel Mobile Internet PC 2004 Platform Vision

Next, and perhaps most importantly, the Gateway Media Center computer, pictured in Figure 21, was designed by Intel's Platform and Architecture Innovations Group, which is a new group within Intel designed to run like IDEO, the much-lauded industrial design firm. As part of the Digital Home initiative, the Media Center PC is currently being sold,

⁶⁶ Extended Mobile Access is technology that enables closed-lid instant access to e-mail and other information through a secondary display on the lid of notebook PCs. Not coincidentally, Insyde Software provided the software that enabled the EMA, including allowing the laptop to enter a low-power mode when the lid is closed, while remaining connected to a wireless enterprise network.

⁶⁷ Intel press release, "Intel Silicon, Platform Innovations Speed Advances In Mobile Computing", February 18, 2004.

but under the Gateway logo. However, the identical product is sold overseas by a different company.⁶⁸



Figure 21: The Gateway Media Center Desktop Designed by Intel

The fact that Intel is now designing desktop and laptop systems is important because it shows that the capability for the vertical integration necessary to deliver a complete system already exists within the company. The more system prototypes (in the case of the Mobile Internet PC 2004 Platform Vision) and actual products (in the case of the Gateway Media Center) that Intel designs, the more they will develop the core competencies, such as industrial design of consumer products, necessary to deliver an entire integrated product, and the more tempted they will be to exploit their new

⁶⁸ Gateway 610 series Media Center product review: http://reviews-zdnet.com.com/Gateway_610_series_Media_Center/4505-3118_16-30565073.html

competencies to “capture the rents” by becoming a vertically integrated proprietary system supplier.

7 Conclusion

The discussion in this thesis has examined the nature of forces acting on the mobile computer industry using a systems thinking approach. First, modularity and integrality of systems and the industries that produce them were explored. Product architecture modularity and industry structure modularity were introduced as useful ways of thinking about the products that an industry makes and way the firms in the industry are organized to accomplish this. Next, the fundamental differences between information and power transfer systems was discussed in the context of Whitney's theory that systems that transmit significant power are inherently more integral. The double-helix framework of Fine and Whitney was then presented, arguing that the organization of firms in an industry cycles between horizontal and vertical due to a number of external and internal pressures, including changes in the modularity or integrality of the industry's product. The second chapter concluded with a historical perspective of Moore's Law and the influence it has had on the computer industry and Intel.

Next, the history of the structure of the computer industry from IISIPA to IISMPA to modular cluster was discussed. It was then hypothesized that changes in the industry brought about by power management difficulties in the mobile platform will cause more integral product architectures. The fundamental physical phenomena of increased power and heat caused by transistor scaling were described next. These side-effects cause the need for better heat management and extended power supply life solutions. These

solutions, in turn, will necessarily become more integral as microprocessor become more like CEMs and the traditional within-module solutions are no longer sufficient to delay the inevitable power crisis.

Chapter 5 then presents a framework to help conceptualize the changes taking place in the computer industry. The axes of product architecture modularity and industry structure modularity form a plane on which any given industry can be plotted. The concept of transaction costs increasing as the product architecture becomes more integral and the industry structure becomes more modular is important to the framework. The computer industry is a point in the modular cluster quadrant of the two-dimensional free body diagram. The trends toward integrality show up as two different forces acting on the industry point mass, both pushing and pulling the product architecture modularity in the more integral direction. The third invisible transaction minimization force acts in the less industry structure modularity and more product architecture modularity direction. This means that the computer industry's resultant trajectory is toward the "IISIPA" quadrant.

These trends toward integrality can be seen at Intel in many ways. The signs of verticality include: chipset designs with more integrated power management functions; Intel's "holistic" approach to power management, including many system-level solutions; leadership and participation in industry consortia to address power concerns; strategic

partnerships and investments by Intel Capital in start-up companies that are developing enabling technology for the next generation power management solutions; and the development of reference and mass market PC products. It is the opinion of the author that it is only a matter of time before significant consolidation is seen within the computer industry due the fundamental modularity constraints that these power management issues cause.

The framework can be a useful systems thinking tool for decision makers at firms in the computer and other industries. For example, it can be used to:

- Decide what strategic investments to make; which skills to invest in and what types of knowledge and capacity are important in the future. (i.e. Should the firm specialize in one component or subsystem or is diversification the best route? For example, given the resultant trajectory vector of Figure 17, Intel should be (and is) investing in other parts of the computer system architecture. This allows them to hedge their investments in the microprocessor component and position themselves to capture market share in a new, vertically integrated industry structure in which value is contributed by more than just the microprocessor component.
- Help indicate the volatility of the industry going forward. (i.e. How quickly have firms responded to the architecture changes in the past?) For example, based on the history of the computer industry given in Section 3.1, it took approximately 5-

10 years for the industry to move from IISIPA to IISMPA after the introduction of the System/360 and the same length of time to go from IISMPA to a modular cluster after the introduction of the IBM PC. Given that the significant integrating force of power management has begun to limit the traditional value metric of processor speed within the last year in mobile computers (with the introduction of Centrino), a similar timeframe can be expected for the transition back to IISIPA. This means that the firms that want to be successful in the new computer industry must position themselves for success by 2015, at the latest.

Further application of the framework that this thesis does not explore includes:

- Forming a strategy to gain a competitive advantage over other firms in the industry. (i.e. purposely try to change the trajectory of an industry by enabling lower or higher transaction costs; this could be done through M&A activity or divestiture to change the modularity of the industry or through new product architectures that are more or less integral) For example, Dell could attempt to protect its position in the industry by decreasing the incentive for Intel and Microsoft to integrate sales and distribution of computer systems by:
 - Keeping transaction costs low for these companies to do business with Dell
 - Enabling new thermal management and battery technologies

- Working with the component suppliers to define better design rules for their respective parts of laptop computers (although this would probably only delay, rather than halt, the transition to IISIPA)

Of course, all of the above are items that decision-makers would ordinarily think about, but this framework gives them a new context in which to frame their line of reasoning for questions such as, “What is the best product architecture?”, “What skills do we invest in or buy?”, and “How do we keep industry dynamics from creating new competitors or rendering our expertise obsolete?”

As Intel develops more and more of the skills necessary to become a vertically integrated supplier of computer systems, they may look to acquire additional skills that are not among their core competencies by merger or acquisition. For example, once all of the design skills are in place to produce an integrated mobile computing solution, Intel may try to acquire supply chain and logistics skills aimed at delivering the systems to end users by buying Dell or Gateway.

Furthermore, this trip back around the double helix may pave the way for a revolutionary new system architecture for computers. Christensen’s work on “disruptive technologies” shows that major architectural innovation usually occurs within a vertically integrated

firm with few suppliers.⁶⁹ Thus, whether the firm is an Intel-Dell merger or another new, yet-unknown giant, once a vertically integrated firm is again able to deliver the entire computer system to the customer, like the IBM of the early 60's that invented the System/360, it could lay the foundation for the next wave of computing (and, if the PC revolution has shown us anything, a change in the way we live our lives).

The complex technological and strategic issues raised in this thesis lead to numerous unanswered questions that could be explored with further research:

- Do any of the generalizations made about product architecture and industry structure apply to other industries besides mobile computing? Or is the computer industry special due being controlled to a large extent by a first-tier supplier? Does the fact that this first-tier supplier's product is an information processing device that is becoming more like a CEM make this a unique dynamic or are there other comparable external forces?
- What does this mean for Dell and Gateway, large PC OEM firms that have only a supply chain competency? Can these firms even begin to acquire the semiconductor design and fabrication skills required of a vertically integrated firm? If not, will these firms be squeezed out of business once a vertically integrated supplier is able to sell directly to customers or will they retain

⁶⁹ Christensen, Clayton M., *The Innovator's Dilemma*, Boston: Harvard Business School Press, 1997.

significant power as to keep the sales and distribution a horizontal portion of the industry?

- Will the mobile computer industry's trajectory split from the desktop computer industry since the forces acting on them are not the same (i.e. the power management issues, although still present, are not as urgent in desktops)? Or are the current mobile computing forces just a precursor for a similar trend in the desktop platform?
- Will the microprocessor become a commodity part in a new, highly integrated mobile computer architecture? If so, can Intel shift their investments and core competencies in time to remain a \$150B+ market capitalization company?

Whitney's concluding remarks in his 1996 paper about information versus power systems were that the current materials used for making microelectronics would "no longer be applicable in two or three process generations (i.e. much beyond 2000)...That is, VLSI may not be like VLSI much longer." Although he was speaking of limitations in VLSI design due to interconnect scaling and not power management, the idea still rings true: that that information processors are starting to look more like CEMs and becoming more integral in their architectures. This will inevitably influence the organization of the industry to become more integral with greater vertical integration within its firms. Indeed, Intel may not be like Intel much longer.

Bibliography

- [Alexander, 2003] Steve Alexander, "Sales of Laptops Zip Ahead of Desktops", Wireless NewsFactor, July 7, 2003, <http://wireless.newsfactor.com/perl/story/21861.html>.
- [Baldwin and Clark, 2000] Baldwin, Carliss Y. and Kim B. Clark, *Design Rules: The Power of Modularity, Volume 1*, The MIT Press, 2000.
- [Bernstein, 2003] Bernstein, Kerry, IBM Thomas J. Watson Research Center, "Caution Flag Out: Microarchitecture's Race for Power Performance", Keynote Address at the 36th International Symposium on Microarchitecture, December 3, 2003.
- [Boutin, 2003] Boutin, Paul, "Stop the Clock: Intel's Centrino shows the real meaning of Moore's Law", *Slate*, March 13, 2003: <http://slate.msn.com/id/2080097/>.
- [Christensen, 1997] Christensen, Clayton M., *The Innovator's Dilemma*, Boston: Harvard Business School Press, 1997.
- [Chu, 2003] Chu, Richard C., "The Perpetual Challenges of Electronics Cooling Technology for Computer Product Applications – from Laptop to Supercomputer", National Taiwan University presentation, sponsored by IBM-Taiwan, November 12, 2003.
- [Clark, 2003] Clark, Don, "Big Bet Behind Intel Comeback: In Chips, Speed Isn't Everything", *The Wall Street Journal*, Vol. 242, No. 99, November 18, 2003.
- [Crawley, 2003] Crawley, Ed, System Architecture class notes, January 2003.
- [Dalziel, 2002] Dalziel, Margaret, "The Impact of Product Architecture Role on Small Photonics and Opto-Electronics Companies in Ottawa", University of Ottawa School of Management Research Project, 2002.
- [Evans, 2003] Evans, Thomas C., "Statistical Usage Models in Mobile Processor Thermal Design and Testing", master's thesis, MIT, June 2003.
- [Fine, 1998] Fine, Charles H., *Clockspeed: Winning Industry Control in the Age of Temporary Advantage*, New York: Perseus Books, 1998.
- [Fine and Whitney, 1996] Fine, Charles H. and Daniel E. Whitney, "Is the Make-Buy Decision Process a Core Competence?", MIT Center for Technology, Policy, and Industrial Development, Cambridge, MA, February 1996.

- [Fujimoto, 2002] Fujimoto, T. "Architecture, Capability, and Competitiveness of Firms and Industries", University of Tokyo Discussion Paper, 2002.
- [Gochman, et al, 2003] Gochman, et al, "The Intel Pentium M Processor: Microarchitecture and Performance", *Intel Technology Journal*, Vol. 7, Issue 2, May 21, 2003.
- [Gray and Meyer, 1993] Gray, Paul R. and Robert G. Meyer, *Analysis and Design of Analog Integrated Circuits*, Third Edition, New York: John Wiley & Sons, 1993.
- [Kolinski, Chary, Henroid, and Press, 2001] Kolinski, Jerzy, Ram Chary, Andrew Henroid, and Barry Press, *Building the Power-Efficient PC: A Developer's Guide to ACPI Power Management*, Intel Press, 2001.
- [Krishnan and Gupta, 2001] Krishnan and Gupta, "Appropriateness and Impact of Platform-Based Product Development", *Management Science*, 2001.
- [Kunii, 2003] Kunii, Irene M., "Japan's Power Play", *Business Week*, October 13, 2003
- [Mahapatra, Parikh, Rao, Viswanathan, and Vasi, 2000] Mahapatra, S., Chetan D. Parikh, V. Ramgopal Rao, Chand R. Viswanathan, and Juzer Vasi, "Device Scaling Effects on Hot-Carrier Induced Interface and Oxide-Trapped Charge Distributions in MOSFET's", *IEEE Transactions on Electron Devices*, Vol. 47, No. 4, April 2000.
- [Meyer and Lehnerd, 1997] Meyer, M.H. and A. P. Lehnerd, *The Power of Product Platforms*, New York: The Free Press, 1997.
- [Moore, 1965] Moore, Gordon, "Cramming More Components Onto Integrated Circuits", *Electronics*, Vol. 38, No. 8, 1965.
- [Murman, et al, 2002] Murman, *et al*, *Lean Enterprise Value*, Palgrave: New York, 2002.
- [Robertson and Ulrich, 1998] Robertson, D. and Karl Ulrich, "Planning for Product Platforms", *Sloan Management Review*, Summer 1998, Vol. 39, No. 4, pp. 19-31.
- [Rogers, 2003] Rogers, Michael, "The Energy Crisis on Your Lap", *Newsweek.com*, November 11, 2003: <http://msnbc.msn.com/id/3475003/>.
- [Rusu, 2001] Rusu, Stefan, "Trends and Challenges in VLSI Technology Scaling Towards 100nm", 34th European Solid-State Device Research Conference, 2001.

- [Salkever, 2003] Salkever, Alex, "Taking It Slow with a Superfast Chip", *BusinessWeek Online*, August 20, 2003:
http://www.businessweek.com/technology/content/aug2003/tc20030820_5175_tc056.htm.
- [Suh, 1990] Suh, Nam P., *The Principles of Design*, Oxford University Press, 1990.
- [Ulrich and Eppinger, 2000] Ulrich, Karl T. and Steven D. Eppinger, *Product Design and Development*, 2nd Ed, McGraw-Hill, 2000.
- [Thompson, Packan, and Bohr, 1998] Thompson, Scott, Paul Packan, and Mark Bohr, "MOS Scaling: Transistor Challenges for the 21st Century", *Intel Technology Journal*, 3rd Quarter 1998.
- [Whitney, 1996] Whitney, Daniel E., "Why Mechanical Design Cannot Be Like VLSI Design", *Research in Engineering Design*, 1996, No. 8, pp. 125-138.