TAPE AUTOMATED BONDING: PRODUCT AND PROCESS RAMP-UP

by

BENJAMIN R. SAMUELS

B.S. Materials Science and Engineering — University of Arizona (1988)

Submitted to the
Department of Materials Science and Engineering and
Department of Management

in partial fulfillment of the degrees of

MASTER OF SCIENCE IN MATERIALS ENGINEERING and
MASTER OF SCIENCE IN MANAGEMENT

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June, 1990

Signature of Author ______________________________
Department of Materials Science and Engineering
Department of Management
May, 1990

Certified by ______________________________
Thomas Eagar
Thesis Supervisor, Department of Materials Science and Engineering

Certified by ______________________________
Marcie J. Tyre
Thesis Supervisor, M.I.T. Sloan School of Management

Accepted by ______________________________
Thomas Eagar
Chairman, Committee on Graduate Students, M.S.E. Department

Accepted by ______________________________
Jeffrey Barks
Associate Dean, M.I.T. Sloan School of Management

TAPE AUTOMATED BONDING:
PRODUCT AND PROCESS RAMP-UP

by

BENJAMIN R. SAMUELS

Submitted to the Department of Materials Science and Engineering
and the Sloan School of Management
in Partial Fulfillment of the Requirements for the Degrees of
Master of Science in Materials Engineering and
Master of Science in Engineering

ABSTRACT

Tape automated bonding (TAB) is currently the fastest-growing electronic packaging technology. It offers significant lead density advantages over conventional wirebonding and significant heat dissipation advantages over flip-chip techniques. Motorola, Inc. and Digital Equipment Corporation selected advanced TAB for the Motorola chip which drives Digital's new entry into the high-performance computer market.

The materials technologies involved in the TAB development efforts are leading edge. This thesis details some of the research performed in support of process development efforts at Motorola's TAB facility. It characterizes yield variation at Motorola and traces it to specific steps in a vendor's TAB tape manufacturing process.

Motorola and Digital are involved in multiple product and process ramp-ups in support of Digital's computer. Management literature is beginning to recognize that the metrics for success in ramp-ups are sometimes very different from those of a mature technology. Companies which compete in products with very short life cycles must learn to manage the problem solving process in environments characterized by high pressure and uncertainty. This thesis discusses problem solving under those conditions as it was observed at Motorola in their TAB ramp-up. It presents new insights into the applicability of traditional cognitive models of problem solving and offers specific suggestions for improving the problem solving environment in the Motorola ramp-up.

ADVISORS:

Prof. Thomas Eagar, Leaders for Manufacturing Professor
Department of Materials Science and Engineering

Prof. Marcie J. Tyre
M.I.T. Sloan School of Management

ACKNOWLEDGEMENTS

I would like to acknowledge M.I.T. and the Leaders for Manufacturing Program for supporting this work and my education. I would also like to thank Motorola and Digital for their confidence in assigning LFM students to a project with real importance to both companies.

I would like to thank my internship apartment/office roommate, Rich Krueger, for not finding any occasion to strangle me in my sleep and for being a great friend. I would also like to thank Julie Johnson and Ginny Maggs for their never-ending enthusiasm and rapt attention as I spewed non-sequiturs at them ad nauseam.

Finally, I would like to thank my family for their moral support throughout my education and allowing me to fly them all to Boston for my graduation.

I would like to apologize for the format of most of the data in this document. The vast majority of the data obtained in support of this thesis has been altered or deleted to protect proprietary information at both Motorola and Digital.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: INTRODUCTION

"Without innovative packaging and interconnection technologies, the revolutionary advances taking place in integrated circuits will be lost. The rapid growth of the integrated circuit industry has tended to hide the fact that not only do all the other components comprise a much larger market value, but packaging and interconnection has now become one of the major barriers to improving both system cost and performance."

M.G. Sage, BPA Ltd., 1987

Conventional wisdom holds that the explosive growth in the microelectronics industry has been based on methodical advances in circuit integration in the chip. But, to paraphrase Sage[1], a microchip is only so much contaminated silicon without complementary interconnection technology linking it to the "outside world."

The most frequently used packaging technology is wirebonding. The "wire" in wirebonding refers to the fine loop of conducting wire which connects the bond pad on the chip to a lead in a lead frame or to a pin in a pin grid array. Highly automated and flexible equipment has made wirebonding the dominant technique for three decades[2].

Unfortunately, wirebonding techniques can not keep pace with increasing lead density. The number of input/output junctions (I/O's) and leads is an exponential function of the number of logic circuits on a chip[*][3]. Increases in circuit density have begun to exceed the capability of conventional wirebonding. Wirebonding tooling fundamentally limits its maximum lead density[4,5]

In addition, the wirebonding lead pitch is also limited by the curvature of the loop which connects the bond pad to the lead frame. If the leads are spaced too closely together, there is a significant risk that the loops might droop slightly and contact their neighbors, causing a short circuit. Foreseeable limitations on wirebonding densities are roughly 300 connections on the periphery of a one sq.-cm. chip. Estimates of maximum number of leads on a 1cm-square chip are on the order of 300 for wirebonding, 1000 for TAB and over ten thousand for flip chip[6,7]. Finally, independent of the lead density considerations, even at 0.2 seconds per bond, wirebonders can become a bottleneck for devices with hundreds of leads.

---

[*] The relationship between the number of logic cells in a chip and the number of signal I/O's is called Rent's Rule. The formula for Rent's Rule is:

$$ckt = (N/K)^n$$

where ckt is the average number of logic cells supported by N number of I/O's. K is a circuit-utilization constant and n is another constant that has been determined through trial and error to be around 1.8.

CHAPTER 1: INTRODUCTION

Motorola, Inc. and Digital Equipment Corporation (DEC) can not use wirebonding for the Hyperchip[*]. DEC uses Hyperchip in its new HiEnd computer[*]. Hyperchip's lead pitch exceeds the capabilities of conventional wirebonding equipment[**]. Though next-generation wirebonding techniques are predicted to be just capable of Hyperchip's present lead density, future configurations of Hyperchip necessitate an entirely different approach to packaging the device. Aside from the purely physical limitations of tooling, wirebonder memory becomes a constraining factor for multichip substrates with complex arrangements of high-I/O chips[8].

There are two basic strategies for packaging chips with high lead densities: controlled collapsed chip connection (C4 or "flip chip") and tape automated bonding (TAB). IBM developed flip chip technology for its mainframe computers. In the flip chip configuration, bond pads are arranged in an area array on the active face of the chip. Tin-lead solder balls are plated on the bond pads. The chip is then connected to a multichip board by placing it active-face-down on the solder balls on the substrate[9,10].

General Electric introduced TAB in 1970 as a highly automated

---

[*] The name of the chip and the computer are disguised to protect proprietary information.

[**] Exact pitch is not provided to protect proprietary information.

alternative to the slow and unreliable wirebonding techniques. But before TAB became widespread, wirebonding advanced to the point where its per-bond unit cost and reliability combined with its inherent flexibility made it superior to GE's new innovation[11]. The complementary chip-bumping technology was expensive and deterred widespread use of TAB[12]. Wirebonding became entrenched as the dominant packaging methodology.

Until the mid-eighties, the only major users of TAB were several Japanese consumer electronics companies which used it for flat-profile products. But in the last decade, TAB has re-emerged as a method for packaging devices with very high lead densities.

Like wirebonding, TAB is a peripheral packaging technique. All of the I/O's, ground and power terminations are located at the edges of the chip as opposed to being distributed across the face of the chip in an area array. With wirebonding, a fine loop of gold or aluminum connects the bond pads to the lead frame, while in TAB, the lead frame connects directly to the bond pads. The TAB lead frame is supported on a mylar (polyimide) tape substrate. The individual leads in the TAB lead frame (or tape site) are patterned using photolithography techniques. Depending on the fabrication process, the leads may be formed by subtractively etching from a common metal film or by plating-up material in channels etched in photoresist. In either scenario, lead pitch is limited by the resolution of the photolithography and not by the bonding equipment. Appendix 1.1 shows that the potential spacing on TAB leads may be half that of the most advanced wirebonding techniques (quad-flat pack)[13].

The primary demand for TAB is for flat profile packages in watches, calculators and LCD's, but an emerging application is for high lead density chips. Such applications include VHSIC (very high speed integrated circuit) devices and custom and semi-custom ASIC's and gate arrays. The worldwide market for TAB products has been estimated to grow at over 10% per year[14]. Because of their familiarity with TAB in low-profile uses, Japanese consumer electronics giants generally have a lead over the rest of the world in TAB technology[15]. By one estimate, the Japanese TAB market will expand by as much as 50% per year for the next several years, though most of this growth is expected in low-profile applications rather than in high-I/O areas[16].

Appendix 1.2 highlights uses of TAB in existing products, ranging from consumer electronics to supercomputers[17]. Many projections for TAB anticipate as many as 1000 leads per cm^2 within the next five years[18].

DEC and Motorola opted for TAB over flip chip for three reasons. First, it is difficult to make electrical contact with all I/O's in a C4 area array because many of them are physically occluded. Only sophisticated customers might have the testing technology for probing C4-mounted chips. Though DEC qualifies as a "sophisticated customer", Motorola wants to eventually sell high-I/O Hyperchip to the broader market. Also, TAB permits "test-on-tape", allowing complete diagnostic evaluation and burn-in prior to mounting on an expensive substrate. Table 1.1 lists basic operations involved in each of the three packaging methodologies. Flip chip and wirebonding do not allow for "test-on-tape"[19]:

TABLE 1.1: PACKAGING ASSEMBLY STEPS

C4	Wirebond	TAB
1. Flux site	1. Solder or epoxy preform place	1. Reel tape into bond position
2. Align and place chip(s)	2. Align and place chip	2. Align and place chip
3. Reflow to bond all pads (On all chips)	3. Die bond	3. Inner-lead-bond (One-chip-at-a-time)
4. Clean flux	4. Wirebond each wire on chip and S/S (One-wire-at-a-time)	4. Encapsulate
5. Test	5. Test	5. Test/burn-in (optional)
6. Encapsulate/ Finish module assembly	6. Encapsulate/ Finish module assembly	6. Excise chip
		7. Align and place assembly on S/S
		8. Outer-lead-bond
		9. Test

Reliability is the second critical issue — especially in a high end computer with many chips. It is difficult to individually perform accelerated life-cycle testing (burn-in) on single C4 chips. The smallest unit that can be easily burned-in is the multichip substrate. It is much more economical to catch flaws at the chip level than at the multichip substrate level when much more value has been added. DEC insisted on burn-in of individual chips. C4 was not as amenable to that requirement as was TAB[20].

Heat dissipation is the final reason for using TAB. Hyperchip dissipates between 20 and 50 watts[21]. Heat management is therefore critical for this device. DEC back-bonds Hyperchip to provide a direct thermal conduction path to the common heat sink mounted on the back of the multichip substrate[22,23].

In flip chip mounting, surface tension holds the chip in place while the solder ball is reflowed to bond it to lands on the substrate. Unlike epoxy bonds, solder-ball bonds are not strong enough to support bulky heat sinks on the back side of the chips during thermal cycling. IBM has developed piston-like heat sinks which are spring-mounted to maintain contact with the die during thermal cycling. These arrangements, however, only dissipate up to four watts per chip[24] - far short of the potential of Hyperchip. Large thermal excursions possible with Hyperchip also exacerbate differences in thermal expansion between the chips and the substrate. Fatigue failure due to temperature cycling is much more likely with solder balls than with a strong adhesive layer.

HYPERCHIP PRODUCTION CHAIN

There are four basic operations in a TAB packaging process. Two are fabrication processes and two are assembly processes. The first is the chip fabrication process, an entirely new technology for Hyperchip. A critical part of the chip process is "bumping". Motorola plates gold bumps on the aluminum bond pads of the chip to provide a stage for bonding the TAB leads[25]. The second fabrication process is the production of the TAB tape. The assembly steps are inner lead bonding (ILB) and outer lead bonding (OLB) during which the tape site is mounted to the die and the entire package mounted to a multi-chip substrate or discrete pin grid array, respectively.

Hyperchip Device

The Hyperchip TAB configuration has more than 250 TAB leads. Hyperchip comes in two basic configurations: low power and high power. The low power array dissipates from 5-15 watts chip. The high power configuration dissipates from 20-50 watts per chip.

The bump process is not formally part of Hyperchip fabrication. The cost of bump facilities was one of the key obstacles to widespread TAB in the 1970's. Motorola first sputters a thin layer of barrier metallization across the surface of the wafer*. The barrier layer prevents the Al of the bond pad from directly contacting the Au of the bump. The purpose is to prevent the formation of brittle Al-Au intermetallics known as "purple plague." The barrier metallurgy also acts as a common bus on which to electroplate the body of the Au bump. Operators plate the gold bumps up from the barrier metal in cavities etched in photoresist. They then strip the photoresist and etch the wafers to remove the barrier metal from between bumps[26,27].

Few customers or tape suppliers have the wafer fabrication technology necessary to plate gold bumps on aluminum bond pads. They rely on Motorola for TAB bumping[28]. Motorola has invested heavily in establishing good bump practices. During their process development, Motorola used only production line equipment. They were able to implement process control in a laboratory setting and ensure that the process remained in control in the production environment[29].

Hyperchip TAB Tape

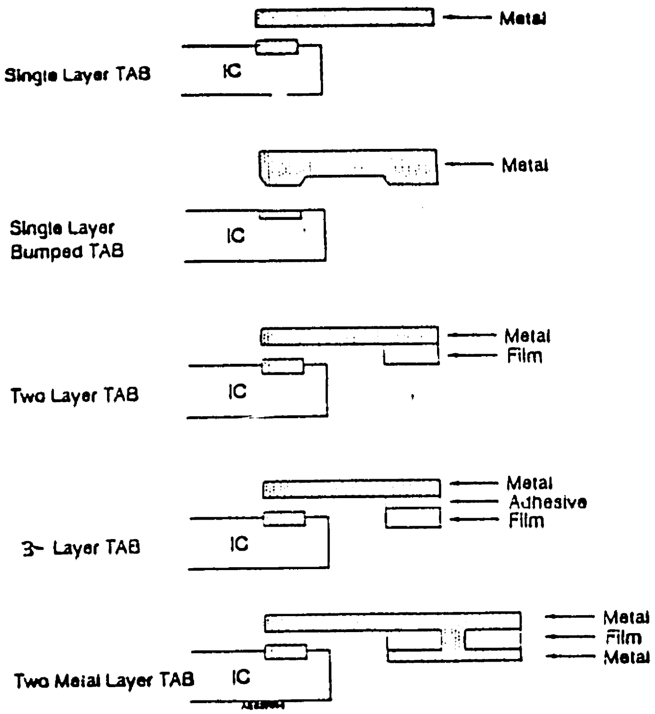
Hyperchip TAB tape itself is a leading-edge technology. The centerline pitch is tighter than most other products currently on the market. Its cross-sectional geometry is also leading edge. Figure 1.1 (following page) shows cross sections typical of tape configurations

---

* The composition of the barrier metallurgy is proprietary. Typical systems used for Au-Al barrier layers have been Cu, Pd, Pt and Ni. See Reference #2, p. 426.

currently in wide-scale production[30].

FIGURE 1.1: Types of TAB Tape Currently in Production



In 1985 when the HiEnd/Hyperchip project was conceived, Motorola and Digital divided the problem into areas of strategic competencies. Motorola concentrated on the fabrication of the chips and the inner lead TAB operations. Digital focussed on developing a fabrication process for its multichip substrate, and on the OLB process. Motorola's new chip fab and bumping technology and DEC's wafer-scale integration techniques for the multichip substrate were extremely complex and highly proprietary. In contrast, the task of developing the TAB tape appeared to be less complicated and less likely to be of strategic value to either company. It seemed to be a natural job for outsourcing to a vendor.

One process for tape-making is to additively plate on a polyimide tape substrate. The various plating solutions at various temperatures stretch and deform the tape slightly in unpredictable directions. This stretching can be a source of yield loss for very high density TAB lead frames[31]. An alternative method is to plate from a fixed substrate. Plating from a fixed substrate eliminates the stretching problem but introduces complications of removing the TAB tape from the substrate at the end fabrication steps.

Inner Lead Bonding Operation

Inner lead bonding is the operation in which Motorola connects the inner leads of the tape site to the bumps on the die. Typical bonding parameters are temperature, force (pressure) and time. Motorola's ILB operation involves forming a eutectic bond between the gold of the "bump" and the tin bonding layer of the inner leads[32]. Motorola uses a tin-gold eutectic rather than a more traditional gold-gold thermocompression bond because the Sn-Au eutectic forms at much lower temperatures than those necessary to thermally activate gold for diffusion bonding. Gold self-diffusion is "fast" only at relatively high temperatures (approaching its melting point). Figure 1.2 shows the dependence of gold self-diffusivity on temperature[*][33].

---

[*] The relationship of the self-diffusion coefficient of gold with temperature is:

$$D = .04 \exp(-170/RT) + .56 \exp(-229/RT)$$

FIGURE 1.2: Temperature-Dependence of Gold Self-Diffusion Coefficient

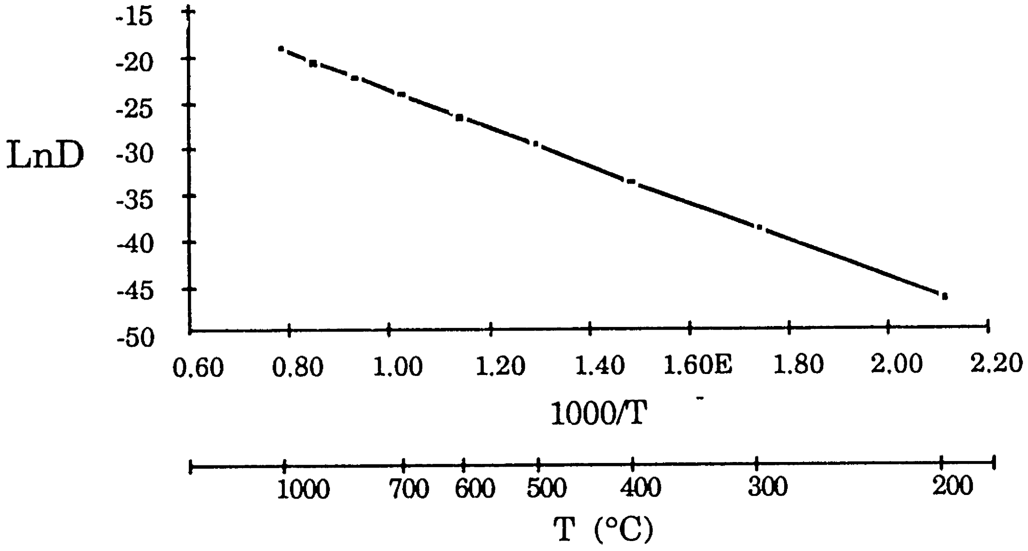
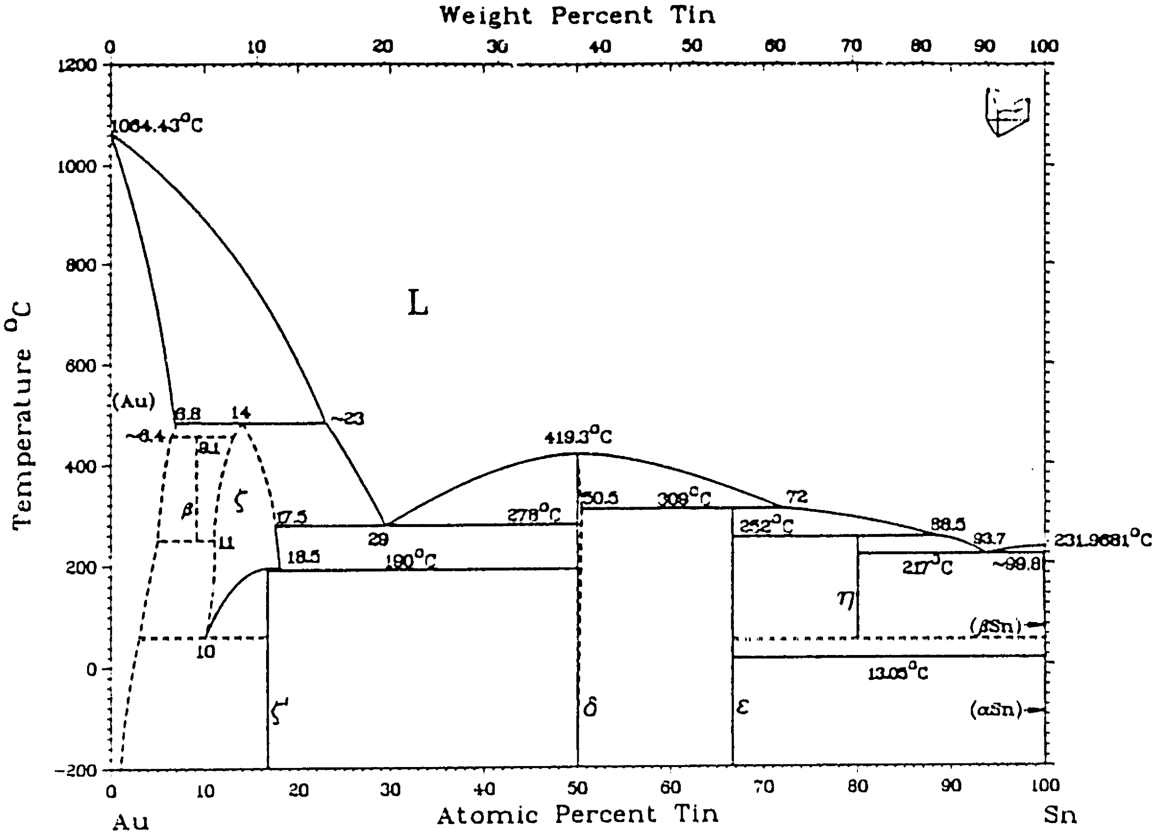


Figure 1.3 is a phase diagram of the Sn-Au binary system[34]. The melting point of gold is 1064°C, far higher than the eutectic temperature of 278°C for an Au_5Sn-AuSn mixture. The melting temperature of tin is only 232°C. The eutectic bonding process may be divided into a series of reactions: 1) Melting and superheating of tin from the bonding layer of the tape; 2) Dissolution of gold from the bump in the molten tin; and, 3) Solidification of a eutectic mixture of Au and Sn upon cooling.

---

where the 170 and 229 have units of kJ/mol, T is expressed in degrees Kelvin and D has units of cm^2/sec.

FIGURE 1.3: Gold-Tin Phase Diagram



There are two schools of thought on bonding TAB leads to the chip. One philosophy is to bond leads one-at-a-time: single-point bonding. The other method is to bond all the leads at once: gang bonding. Gang bonding offers significant throughput advantages over single-point bonding, especially for tape sites with hundreds of leads. Motorola has opted for gang bonding, anticipating a day when their TAB operation will be a high volume production line. The thermode applies the temperature and pressure necessary to drive the eutectic reaction[35].

A critical factor in gang bonding is co-planarity of the thermode with the die. With an uneven thermode, in some areas of the die, the silicon may be damaged by excessive pressure while in other regions, leads may remain unbonded due to lack of pressure. Operators periodically re-align them with the thermode. After multiple uses, thermode planarity begins to degrade due to bowing from temperature cycling, oxidation and contamination from the metals on the TAB lead frame. Motorola re-surfaces their thermodes on a regular basis. An even temperature distribution across the bonding surface of the thermode is also critical. With uneven temperature distribution, some areas may be damaged by excessive heat while others remain unbonded. Finally, the cleanliness of the bonding surfaces is also critical. In wirebonding, ultrasonic vibration is often applied along with temperature (thermosonic bonding) which breaks up oxides and other contaminants on the surface. In gang bonding, there is no scrubbing effect. To ensure a reliable bond at an acceptably low temperature and pressure, surfaces must be clean prior to gang bonding and the temperature must be sufficient to cause the three-step eutectic reaction described above[36].

Outer Lead Bonding and DEC Technologies

The niche for the HiEnd is for current VAX users who wish to trade-up and expand on their established VAX base. Previously, these users had to switch to a competitor's mainframe if they wanted to increase their performance. The price tag for the HiEnd will be a fraction of the cost of competitors' high-end computers of similar capacity. Finally, like all new DEC computers, it is vertically compatible with all existing DEC machines by virtue of DEC's common operating system, VMS.

The HiEnd gets all this power from processor(s) housing multiple multichip substrates. Each substrate holds multiple Hyperchips, static RAM's and timing devices. Chip-to-chip wiring is through metallization in laminates of the substrate.

ENGINEERING PROBLEM STATEMENT

In 1988, Motorola won the first annual Malcolm Baldridge Award. Two philosophies permeate Motorola's entire organization: Total Customer Satisfaction and "Six Sigma". Total Customer Satisfaction is Motorola's corporate promise of quality products and service to all of its clientele. Six Sigma is an ideal that plus or minus six standard deviations of product attributes should fall within specification limits. That number translates into a product defect rate of three parts per million. Motorola arrived at this number by competitively benchmarking the products and processes of companies they considered to be world class. They found that six sigma was already attainable for many of those companies. The only way Motorola felt it could compete in the long term was to achieve the same level of process control. The corporate goal is to achieve six sigma across the company by 1992.

"Six-sigma"-type quality is crucial to HiEnd. Although HiEnd is designed to be "fault-tolerant", the machine is so complex that the cumulative effect of marginal quality in its components could be disastrous. Failure in a small fraction of the total number of inner lead bonds could cause the entire computer to fail. The engineering portion of this thesis details some of efforts made to improve the inner lead bond yield of Motorola's TAB tape.

LAYOUT OF ENGINEERING THESIS

Chapter 2 presents an investigation in the variation in a final characteristic of the TAB tape with specific process steps. For the sake of discussion, the correlation of lead thickness with leveling agent was investigated. Two leveling agents, L_1 and L_2 were evaluated for their influence on lead thickness variation*. Chapter 3 discusses a corresponding correlation in ILB yield loss. It concludes that lead thickness variation may combine with other process control problems negatively impact on ILB yield.

---

* The correlation of lead thickness with leveling agent was not ever studied. It is a convenient relationship to discuss because its influence is fairly well characterized on plating literature. It is a realistic situation for any plating operation, not just plating TAB lead frames.

CHAPTER 1: INTRODUCTION

HYPERCHIP DECISION-MAKING ENVIRONMENT

Motorola and DEC have formed a strategic partnership in the Hyperchip program. Buy-in from senior levels of management at both companies complements a broad base of both formal and informal communication between engineers. Daily meetings and conference calls keep parties from both companies informed of latest developments. Both companies have commitments from top executives of critical vendors for cooperation in an intense product and process ramp-up.

However, the immediate needs of firefighting in a ramp-up complicate decision making in the Hyperchip project. Even with open communications, the priorities of one organization are often at odds with those of another. Individual groups meeting local optima lose sight of overall system needs. With as many organizations as have responsibility for Hyperchip, many decisions require networking and buy in of several different organizations. It is often difficult to get this buy in a timely fashion when faced with the technical, budgetary, geographical and cultural constraints which characterize the Hyperchip development effort.

With so many parties involved, the "Not Invented Here" syndrome can become a serious barrier to improvement. In a high-pressure atmosphere characterized by uncertainty, it is difficult not to fall into the pattern of accepting only those solutions which are developed within one's own organization. The Hyperchip project is of paramount importance to both Motorola and DEC. Since Hyperchip TAB is on the critical path of HiEnd, either company would make their extended analytical facilities and engineering talent available to solve the local problems of the TAB development effort. Distrust of externally developed practices and limited time to verify these solutions have limited the Project's utilization of these external resources.

Time pressures are enormous. Meeting project milestones is always problematic. Financial pressure is tremendous. The Hyperchip TAB project and HiEnd represent a tremendous development effort for both Motorola and DEC. The core technologies are seen as platforms for future products, so their success is of paramount importance.

CHAPTER 1: INTRODUCTION

MANAGEMENT PROBLEM STATEMENT

The technical challenges in Hyperchip are enormous. In many respects, however, the managerial challenges are as significant as the engineering issues. This situation is typical of a rapid product and process ramp-up. It has been suggested that the chief task of companies which compete in markets with increasingly short life cycles is not to specialize in any one transitory technology, but to learn the art of managing ramp-ups efficiently[37,38,39].

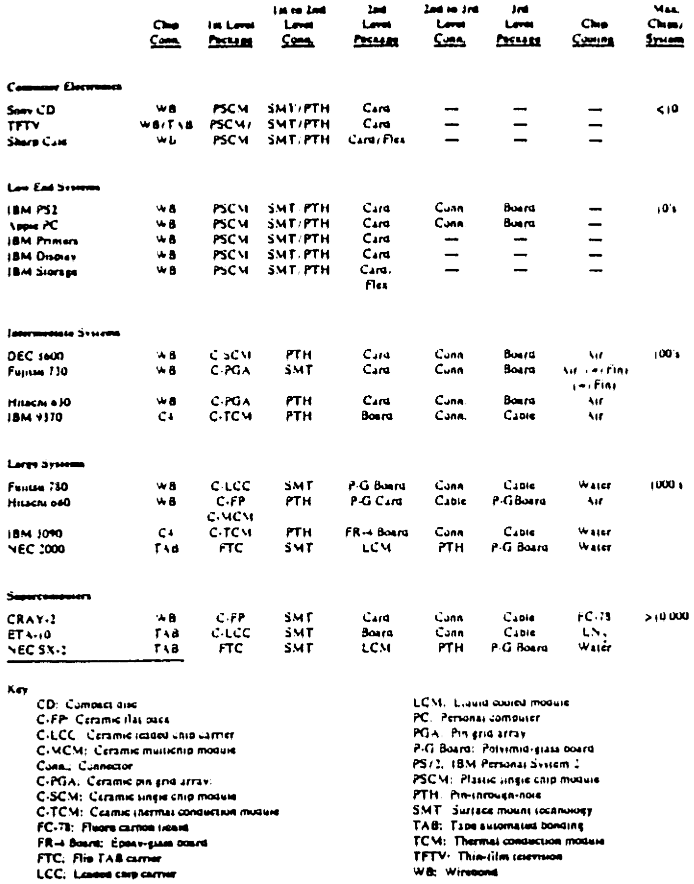
The managerial portion of this thesis addresses the dynamics of decision making under the conditions outlined above. Chapter 4 presents an overview of cognitive models of decision-making. Chapter 5 describes the decision-making environment at Motorola by way of examples that typified the Hyperchip ramp-up. Chapter 6 discusses new insights into the cognitive models presented in Chapter 4 based on the examples in Chapter 5. It suggests ways to change managerial practices to improve decision making in the ramp-up environment.

APPENDIX 1.1: Lead Spacings of Wirebonding and TAB

Package	Package Materials	Minimum Number of Interconnections Available 1985	Minimum Number of Interconnections Future ≈1995	I/O Spacing mm
Dual-in-Line (DIP)	Alumina Ceramic, Plastic	64		2.54
	Plastic	64		2.54
Shrink DIP	Plastic	64		1.77
Skinny DIP	Plastic	64		2.54
Single-In-Line (SIP)	Plastic	21		2.54
Leadless Chip Carrier (LLCC)	Ceramic	132		1.27
		300	400	0.63
	Plastic	180		1.00
Small Outline Package (SOP)	Plastic	40		1.27
Leaded Chip Carrier (PLCC)	Plastic	84		1.27
		144		0.63
Quad Flat Pack (QFP)	Plastic	130		1.00
		160		0.65
	Ceramic	130		0.40
		200		0.63
			400	0.40
Pin-Grid Array (PGA)	Alumina (Single Chip)	312		2.54
			>400	1.27
	Ceramic (Multichip)	2177	>3000	2.54
	SiC	240		2.54
	Plastic	208	>300	2.54
Tape Automated Bonding (TAB)	Plastic	300		0.50
			>400	0.25

CHAPTER 1: INTRODUCTION

APPENDIX 1.2: Applications of Wirebonding, TAB and Flip Chip



	Chip Conn.	1st Level Package	1st to 2nd Level Conn.	2nd Level Package	2nd to 3rd Level Conn.	3rd Level Package	Chip Cooling	Max. Chips/System
Consumer Electronics								
Sony CD	WB	PSCM	SMT/PTH	Card	—	—	—	<10
TFTV	WB/TAB	PSCM/	SMT/PTH	Card	—	—	—	
Sharp Calc	WB	PSCM	SMT/PTH	Card/Flex	—	—	—	
Low End Systems								
IBM PS2	WB	PSCM	SMT/PTH	Card	Conn	Board	—	10's
Apple PC	WB	PSCM	SMT/PTH	Card	Conn.	Board	—	
IBM Printers	WB	PSCM	SMT/PTH	Card	—	—	—	
IBM Display	WB	PSCM	SMT/PTH	Card	—	—	—	
IBM Storage	WB	PSCM	SMT/PTH	Card, Flex	—	—	—	
Intermediate Systems								
DEC 8600	WB	C-SCM	PTH	Card	Conn	Board	Air	100's
Fujitsu 730	WB	C-PGA	SMT	Card	Conn	Board	Air (w/ Fins) (w/ Fins)	
Hitachi 630	WB	C-PGA	PTH	Card	Conn.	Board	Air	
IBM 9370	C4	C-TCM	PTH	Board	Conn.	Cable	Air	
Large Systems								
Fujitsu 780	WB	C-LCC	SMT	P-G Board	Conn	Cable	Water	1000's
Hitachi 680	WB	C-FP C-MCM	PTH	P-G Card	Cable	P-G Board	Air	
IBM 3090	C4	C-TCM	PTH	FR-4 Board	Conn	Cable	Water	
NEC 2000	TAB	FTC	SMT	LCM	PTH	P-G Board	Water	
Supercomputers								
CRAY-2	WB	C-FP	SMT	Card	Conn	Cable	FC-75	>10 000
ETA-10	TAB	C-LCC	SMT	Board	Conn	Cable	LN_2	
NEC SX-2	TAB	FTC	SMT	LCM	PTH	P-G Board	Water	

Key

CD: Compact disc
C-FP: Ceramic flat pack
C-LCC: Ceramic leaded chip carrier
C-MCM: Ceramic multichip module
Conn.: Connector
C-PGA: Ceramic pin grid array
C-SCM: Ceramic single chip module
C-TCM: Ceramic thermal conduction module
FC-75: Fluorocarbon liquid
FR-4 Board: Epoxy-glass board
FTC: Flip TAB carrier
LCC: Leaded chip carrier
LCM: Liquid cooled module
PC: Personal computer
PGA: Pin grid array
P-G Board: Polyimid-glass board
PS/2: IBM Personal System 2
PSCM: Plastic single chip module
PTH: Pin-through-hole
SMT: Surface mount technology
TAB: Tape automated bonding
TCM: Thermal conduction module
TFTV: Thin-film television
WB: Wirebond

CHAPTER 1: INTRODUCTION

1. M. G. Sage, BPA Technology and Management, Ltd, Dorking, Surrey, England, "The Future for Electronic Packaging and Interconnection," paper presented at NEPCON West, 1987

2. N. Koopman, T. Reiley, P. Totta, "Chip-to-Package Interconnections," Chapter 6 of Microelectronics Packaging Handbook, Ed. R. Tummala and E. Rymaszewski, p. 365, 1989,Van Nostrand Reinhold, New York.

3. E. Rymaszewski and R. Tummala, Microelectronics Packaging Handbook, Ed. R. Tummala and E. Rymaszewski, §1.21, p. 13, 1989,Van Nostrand Reinhold, New York; also, W. Heller and W. Mikhail, §2.21 of same text.

4. See Figure 6-31 of Microelectronics Packaging Handbook, Ed. R. Tummala and E. Rymaszewski, p. 400, 1989,Van Nostrand Reinhold, New York.

5. G. Dehaine, K. Kurzweil, and P. Lewandowski. "Crossing the 200 Wire Barrier in VLSI Bonding," *Proc. ISHM Symp.*, p. 353-356, Sept, 1984

6. See Figure 1-16 of Microelectronics Packaging Handbook, Ed. R. Tummala and E. Rymaszewski, p. 30, 1989,Van Nostrand Reinhold, New York.

7. N. Koopman, T. Reiley, and P. Totta. "High Density Chip Interconnections," *ISHM*, Oct., 1988.

8. N. Koopman, T. Reiley, P. Totta, "Chip-to-Package Interconnections," §6.4.6 of Microelectronics Packaging Handbook, Ed. R. Tummala and E. Rymaszewski, p. 402, 1989,Van Nostrand Reinhold, New York.

9. See Figure 6-6 of Microelectronics Packaging Handbook, Ed. R. Tummala and E. Rymaszewski, p. 368, 1989,Van Nostrand Reinhold, New York.

10. L. Miller. "Controlled Collapse Reflow Chip Joining," *IBM J. Res. Develop.*, 13: p. 239-250, May, 1969.

11. Personal communication with Motorola TAB development engineers. June, 1989.

12. N. Koopman, T. Reiley, P. Totta, "Chip-to-Package Interconnections," Chapter 6 of Microelectronics Packaging Handbook, Ed. R. Tummala and E. Rymaszewski, p. 365, 1989,Van Nostrand Reinhold, New York.

13. Table from Microelectronics Packaging Handbook, Ed. R. Tummala and E. Rymaszewski, Table 1-2, p. 26, 1989,Van Nostrand Reinhold, New

York.

14. L. Jardine. "Worldwide Tape Automated Bonding Marketplace," *Proc. of NEPCON West*, Vol 1, p. 119, 1987.

15. W. Beckenbaugh. "TAB technology in Japan," Internal findings of Motorola corporate task force on Japanese TAB, presented at *Fall Symposium of Advanced Manufacturing Technology Council*, Oct. 11, 1989, Scottsdale, AZ.

16. J. Vardaman. "Recent Japanese Developments in Tape Automated Bonding," Empire Planning Report, Ed. by P. Tschirhart-Spangler, p. 6, 1989.

17. Table from Microelectronics Packaging Handbook, Ed. R. Tummala and E. Rymaszewski, Table 1-2, p. 26, 1989,Van Nostrand Reinhold, New York.

18. J. Vardaman. "Recent Japanese Developments in Tape Automated Bonding," Empire Planning Report, Ed. by P. Tschirhart-Spangler, Executive Summary, 1989.

19. Table from Microelectronics Packaging Handbook, Ed. R. Tummala and E. Rymaszewski, Table 6-9, p. 445, 1989, Van Nostrand Reinhold, New York.

20. Private communications with Motorola and Digital Equipment Corporation development engineers.

21. Motorola Hyperchip Design Manual — actual name of reference stricken to protect proprietary information.

22. Private communication with DEC HiEnd project engineers, October, 1989.

23. For typical TAB chip-on-board configurations, see Figure 6-50 of Microelectronics Packaging Handbook, Ed. R. Tummala and E. Rymaszewski, p. 422, 1989, Van Nostrand Reinhold, New York.

24. N. Koopman, T. Reiley, P. Totta, "Chip-to-Package Interconnections," Chapter 6 of Microelectronics Packaging Handbook, Ed. R. Tummala and E. Rymaszewski, p. 390, 1989,Van Nostrand Reinhold, New York.

25. N. Koopman, T. Reiley, P. Totta, "Chip-to-Package Interconnections," Chapter 6 of Microelectronics Packaging Handbook, Ed. R. Tummala and E. Rymaszewski, p. 426, 1989,Van Nostrand Reinhold, New York.

26. Paraphrased from proprietary Motorola bump processing specification last updated Aug., 1989. For schematic of steps in typical bump process,

see Figure 6-54 of Microelectronics Packaging Handbook, Ed. R. Tummala and E. Rymaszewski, p. 427, 1989,Van Nostrand Reinhold, New York.

27. R. Oswald and W. deMiranda. "Application of Tape Automated Bonding Technology to Hybrid Microcircuits," *Solid State Tech.*, p. 33-38, 1977

28. Typical of TAB bumping houses. J. Vardaman. "Recent Japanese Developments in Tape Automated Bonding," Empire Planning Report, Ed. by P. Tschirhart-Spangler, p. 6, 1989.

29. Private communication with Motorola bump process engineers, June, 1989.

30. Figure taken from BPA Consulting report commissioned by Motorola, June, 1989. The original source of figure is not known.

31. Private communications with Motorola and DEC development engineers on HiEnd supply chain, June, 1989.

32. N. Koopman, T. Reiley, P. Totta, "Chip-to-Package Interconnections," Chapter 6 of Microelectronics Packaging Handbook, Ed. R. Tummala and E. Rymaszewski, p. 426, 1989,Van Nostrand Reinhold, New York.

33. P. Shewmon. Diffusion in Solids, 2nd Edition, p. 92, Table 2-6, Pub. by The Minerals, Metals & Materials Society, Warrendale, PA, 1989.

34. Figure from Binary Alloy Phase Diagrams, p. 315-317

35. See Figures 6-39 and 6-55 of N. Koopman, T. Reiley, P. Totta, "Chip-to-Package Interconnections," Chapter 6 of Microelectronics Packaging Handbook, Ed. R. Tummala and E. Rymaszewski, pgs. 411 and 428, 1989,Van Nostrand Reinhold, New York.

36. N. Koopman, T. Reiley, P. Totta, "Chip-to-Package Interconnections," Chapter 6 of Microelectronics Packaging Handbook, Ed. R. Tummala and E. Rymaszewski, 1989,Van Nostrand Reinhold, New York.

37. R. Bohn. "Management of Learning in Ramp-Ups: Research Summary," Harvard Business School Working Paper, April, 1987.

38. R. Bohn. "Learning by Experimentation in Manufacturing," Harvard Business School Working Paper, May, 1987.

39. R. Bohn. "Noise and Learning in Semiconductor Manufacturing," Harvard Business School Working Paper, May, 1988.

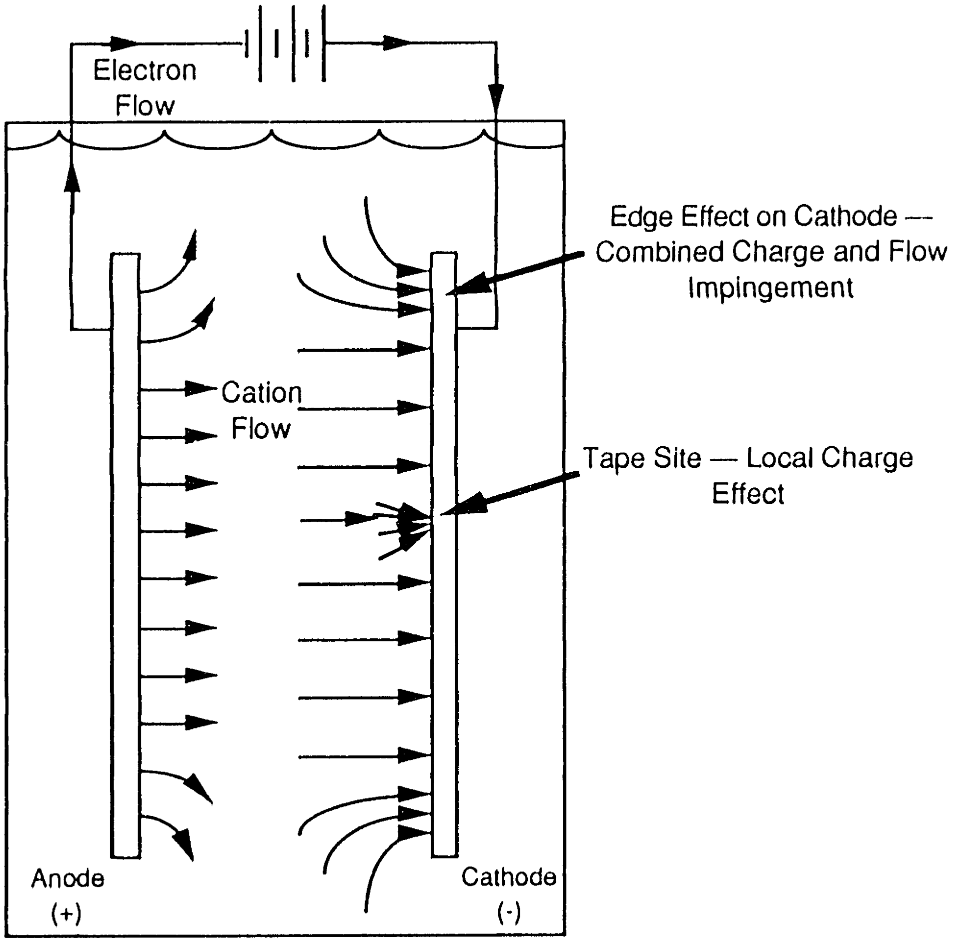
CHAPTER 2: CHARACTERIZING LEAD THICKNESS VARIATION

This chapter summarizes the investigation of lead thickness variation of Motorola TAB tape which has been plated from a fixed substrate. Because this tape is plated from a fixed substrate, it has very good ILB dimensional yield within the plane of the tape. However, periodic differences in bonding parameters required to get acceptable inner lead bonding have pointed to variation the thickness of the leads, in the direction normal to the plane of the tape. The ease of bonding an individual tape site appears to relate to the original process used to plate the tape. In particular ILB yield seems to be related to L_1 and L_2 (two leveling agents used in an extended engineering experiment). Anecdotal evidence from TAB Assembly technicians indicates that tape sites made using L_2 have been easier to bond and have yielded higher on subsequent lead pull tests than have tape sites made using L_1 .

It is hypothesized that yield properties which correlate to L_1 and L_2 stem from their influence on smoothing differences in interface energy during plating. One positive indication of the effectiveness of the leveling agent would be a difference in lead thickness variation[1]. Chapter 2 focusses on demonstrating that inner lead thickness varies in a predictable fashion depending on the leveling agent used. Chapter 3 follows up on this investigation by demonstrating that ILB yield correlates with this variation.

In an electroplating operation, there are predictable variations in current density across the area of a substrate and at unusual geometries on the face of the substrate. These differences stem from the fact that at edges and in corners there are more angles of impingement than there are along flat faces and similarly for unusual geometries on the face of the substrate. Figure 2.1 illustrates this point and shows flow lines of ion transport from the anode to the substrate (cathode):

FIGURE 2.1: Cathode Edge Effects



CHAPTER 2: CHARACTERIZING LEAD THICKNESS VARIATION

In a situation where ionic flow is the rate limiting step, areas of high current density should plate thicker leads than areas of low current density. Thus, the ultimate influence of current density on lead thickness depends on whether or not ionic transport across the electrolyte is the kinetically rate-limiting step. Other possible rate limiting steps are bulk agitation, bulk diffusion in the electrolyte (very small overpotential, little convection), diffusion across the electrical double layer present at the plating surface, and the attachment rate at the surface once ionic material diffuses through the double layer. Leveling agents work by effectively making the rate of attachment at the surface the rate limiting step. They thereby smooth the effect of current density, overpotential, and bulk agitation on lead thickness variation[2].

2.1) INNER LEAD PROFILE (1)

2.1.1) Method*

Multiple engineering lots were removed from the TAB fabrication operation just after plating the bulk of the copper leads. Fourteen lots were measure for average lead thickness using a Dektak 3030 profilometer. These sites were profiled across the inner lead portion of the TAB leads at right angles to the lead approximately 0.5 mm from the tip of the leads.

---

* Descriptions of experimental methods in this document have been changed to protect proprietary information.

2.1.2) Results[*]

The average lead thickness per tape site and the difference between the thickest and thinnest leads in each tape site are tabulated in Table 2.1:

TABLE 2.1: Inner Lead Thickness Dependence on Leveling Agent (1)

Lot No.	Avg. (mil)	StDev. (mil)	Level Agent	Lot No.	Avg. (mil)	StDev. (mil)	Level Agent
1	1.27	0.21	L1	8	0.88	0.13	L2
2	1.10	0.19	L1	9	1.11	0.22	L1
3	1.29	0.29	L1	10	0.90	0.11	L2
4	0.88	0.10	L2	11	0.89	0.07	L2
5	0.88	0.10	L2	12	1.28	0.14	L1
6	0.86	0.16	L2	13	1.12	0.23	L1
7	0.93	0.09	L2	14	1.26	0.25	L1

The data collected in this survey indicate that, in general, leads in lots manufactured using L_1 tend to be thicker on average and have more variation in thickness than leads made using L_2.

2.2) INNER LEAD PROFILE (2)

Several engineers expressed reservations concerning the specific plating solution used in the first engineering experiment. To confirm the results, a second set of runs was conducted.

2.2.1) Method

Fourteen new engineering lots were selected from seven different days. Each tape site was from a different lot to mitigate the influence of

---

[*] Actual numbers presented in results have been changed in this document to protect proprietary information.

fluctuations in plating solution. The total period of the experiment was kept to one week to minimize age effects. The method for measuring lead thickness was the same used in §2.1.

2.2.2) Results

The trends shown in Table 2.2 correspond closely to those in Table 2.1: average lead thickness and the variation in thickness tended to be greater for L_1 than for L_2:

TABLE 2.2: Inner Lead Thickness Dependence on Leveling Agent(2)

Lot No.	Avg. (mils)	StDev. (mils)	Level Agent	Lot No.	Avg. (mils)	StDev. (mils)	Level Agent
1	1.24	0.32	L1	8	1.13	0.17	L1
2	1.22	0.20	L1	9	0.91	0.11	L2
3	1.27	0.22	L1	10	1.13	0.19	L1
4	0.91	0.13	L2	11	0.92	0.12	L2
5	0.92	0.09	L2	12	0.85	0.10	L2
6	0.90	0.07	L2	13	1.10	0.18	L1
7	0.88	0.11	L2	14	1.17	0.21	L1

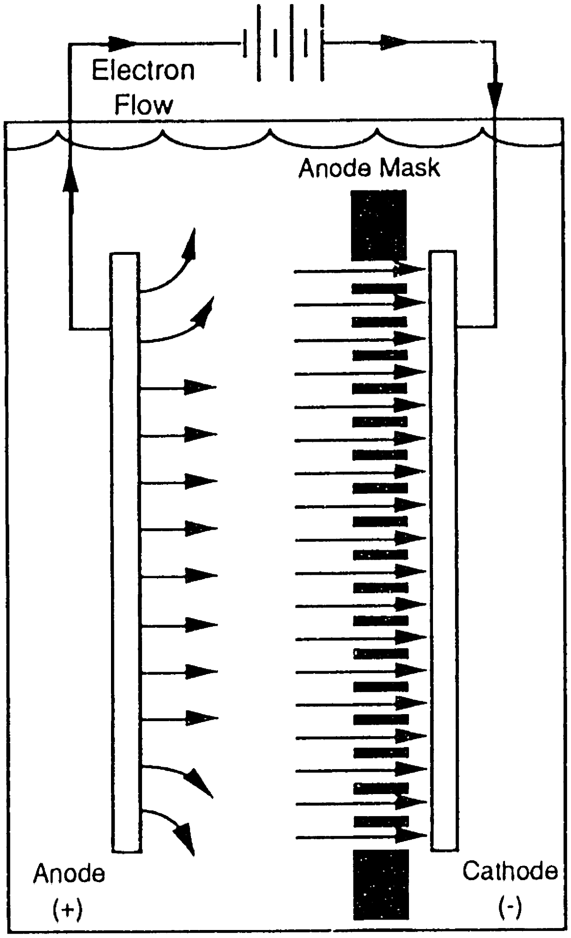
2.3) SUMMARY

2.3.1) DISCUSSION

The results of the lead thickness surveys confirm the hypothesis that lead thickness varies with leveling agent chosen. In general, leads are thicker using L_1 than they are using L_2. These observations suggest that L_2 present more of a rate-limiting step than L_1. It therefore tends to damp-out the current density effects which would result in leads thicker in the corners of the plating substrate than in the center. Figure 2.1 shows how the net arrival of ionic material to the edges of the cathode tends to be greater than at the center.

Two additional solutions have been proposed to minimize current density non-uniformity. One way to control large-scale non-uniformity is to place a physical mask between the anode and the cathode. By carefully selecting the geometry of an anode mask, it is possible to reduce the edge effect on the substrate. Figure 2.3 shows a schematic example on the following page.

FIGURE 2.2: Effect of Anode Mask on Ion Flow Lines



The way to locally alter the charge concentration at the level of the tape site is by tailoring the geometry of current robbers. Corner leads tend to plate thicker than mid-quadrant leads on individual tape sites. The reason is that corners are uniquely exposed. Two leads in the middle of the quadrant "see" the same environment — a lead on either side and about the same current density. Leads at the corners "see" only one other lead or only a few other leads to one side. This problem is exacerbated at the edges of the cathode. There are more angles of impingement at the corners than there are in the middle of the quadrant. There are more angles of impingement at the corners than there are in the middle of the quadrant. The object is to make every lead look electrochemically identical. One way to meet this need is to plate dummy leads and extra material on either side of the corner leads. They will then "see" leads on both sides instead of just on one side. Similarly, it is possible to plate material just off the ends of all the leads so that the tips of the leads appear electrochemically identical to the body of the leads (other investigations have shown that the tips of the leads tend to be thicker than the bulk of the leads).

As of this writing, the current robber geometry has, in fact, been modified in engineering runs to minimize local charge imbalances. Extra material has been placed all around the border of the cathode. Dummy leads have been added next to the corner leads in individual tape sites and mirror image leads have been plated opposite the ends of inner leads. These efforts have met with partial success. The difference in lead thicknesses between the interior and edge sites is less with the experimental geometry than without it. But, the experimental inner lead robber geometry actually appears to increase intra-site thickness variation. The experimental configuration has less metal and less overall robbing effect. Expected benefits of mirror image structures have not emerged partly because the amount robber metal has been reduced and partly because the tape Fab has to leave a significant amount of space between the end of the lead and the robber for the micrometer they use to measure lead thickness.

2.3.2) AREAS FOR FURTHER STUDY

All of the insights into current density generated by this investigation are indirect — based on observations of lead thickness. One avenue for future study is to focus on directly measuring current density in plating tanks with different anode mask and current robber configurations.

Future lead thickness surveys at Motorola can be much more thorough and rapid as new laser measuring equipment comes on-line. They will also be non-destructive — unlike the DEKTAK 3030 which requires sectioning cathodes or bonding parts. The laser equipment will also allow in-process monitoring of lead thickness at various stages of tape manufacturing and assembly. It may obviate the need for the micrometer in in-process lead thickness measurements and allow the tape Fab to plate additional current robber material closer to the ends of the leads in future engineering batches.

Other variables which have not been directly studied are varying agitation during plating and using pulse plating as ways to minimize current density effects. Agitation is a common technique for smoothing plating deposits — especially for bulk recovery of metals in mining operations. There are three transfer mechanisms across the aqueous medium in a plating operation: diffusion, ionic migration, and bulk convection. Under the influence of an electrical bias, in the absence of bulk convection, ionic migration dominates over diffusion. This situation is the one shown in the figures above. Under agitation, bulk motion of the aqueous medium dominates over ionic migration, thereby directly impacting on cathode current density effects[3]. Agitation also minimizes the thickness of the electrical double layer (boundary layer) at the surface of the cathodes, thereby impacting significantly on local charge concentration effects as well as on bulk transport effects[4].

As stated above, in the absence of significant bulk convection, ionic migration dominates. Under conditions of either constant current or constant voltage, edge-thickness build-up is fairly common, as shown in Figure 2.1. The results of Chapter 2 are typical of these scenarios. One technique that has been developed to minimize this effect is pulse plating[5]. Pulse plating is therefore one avenue of investigation with potentially high returns.

2.3.3) CONCLUSIONS

(1) TAB processing cathodes exhibit lead thickness variation correlated with location on the cathode. This variation is mitigated by leveling agents. L_2 appears to have more "leveling" effect than L_1.

(2) Lead thickness also correlates with lead position within

individual tape sites. Leads tend to be thickest in the corners of the quadrants of the individual tape sites than in the centers of the quadrants. Leveling agents also address this problem.

(3) Several existing practices directly address the current density variation suspected in the TAB tape Fab. They should be studied for their implementation in the plating process.

1. E. H. Lyons, Jr.. "Fundamental Principles," p. 28-29 of Modern Electroplating, 3rd Edition, Ed. by F. Lowenheim, 1974, Pub. John Wiley & Sons, New York.

2. V. A. Ettel. "Module 3: Kinetics," of Unit Processes in Extractive Metallurgy, Ed. by L. G. Twidwell, T. J. O'Keefe, V. A. Ettel, and J. D. Robinson, Pub. by National Science Foundation, Grant # SED75-04821

3. V. A. Ettel. "Module 3: Kinetics," of Unit Processes in Extractive Metallurgy, Ed. by L. G. Twidwell, T. J. O'Keefe, V. A. Ettel, and J. D. Robinson, Pub. by National Science Foundation, Grant # SED75-04821

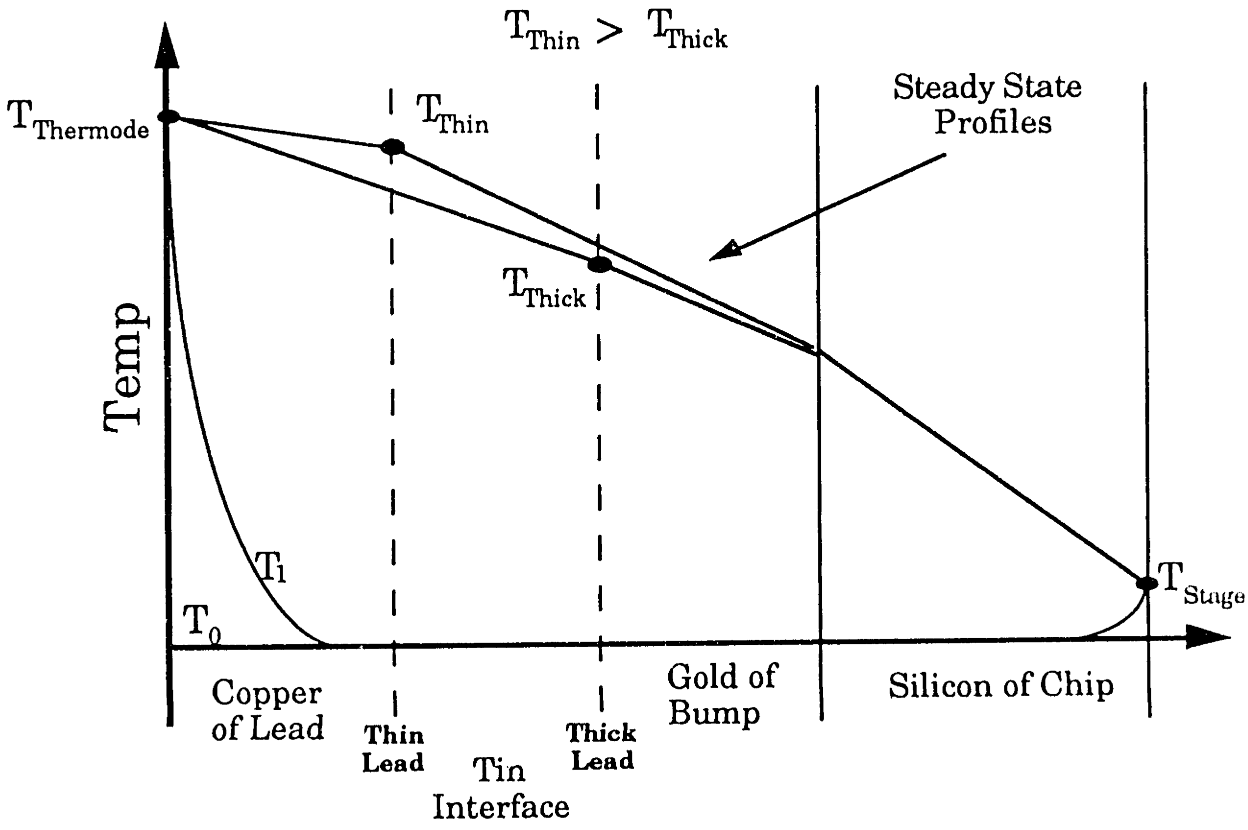
4. "Module 2: Thermodynamics," of Unit Processes in Extractive Metallurgy, page 2.2.15, Ed. by L. G. Twidwell, T. J. O'Keefe, V. A. Ettel, and J. D. Robinson, Pub. by National Science Foundation, Grant # SED75-04821

5. P. Bindra, S. Levine, W. Pimbley, R. Schaffer, J. Reid, W. Underkofler. and R. Galasko. "Metal Deposition," Chapter 16 of Principles of Electronic Packaging, Ed. by D. Seraphim, R. Lasky and C.Y. Li. Publ. McGraw-Hill, New York, 1989.

CHAPTER 3: ILB Yield Studies

Lead thickness variation appears to play a role in ILB yield loss for two reasons. Thick leads provide a longer thermal conduction path from the thermode to the bonding surface of the lead. There are two sources of heat during inner lead bonding: heat from the thermode and heat from the stage at the back side of the die (the die is heated to minimize thermal shock during bonding). The die-temperature is generally much lower than the bonding temperature. The temperature profile from the thermode to the back of the die can be predicted by modeling the thermal conduction path as a series composite wall[1]. Figure 3.1 shows a schematic of the temperature profile:

FIGURE 3.1: Bonding Temperature Profile

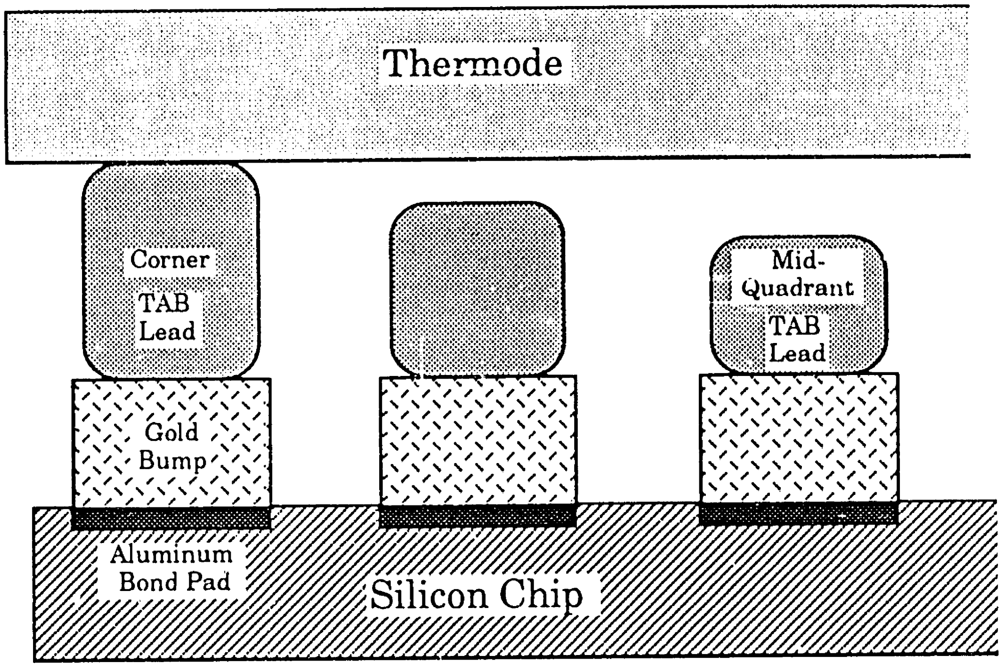


The top line shows the temperature profile at steady state. The curves below are temperatures during heating before steady state is reached. Tin on a thick lead is exposed to lower temperature (for less time) than tin on a thin lead. The result should be poorer bonding in areas with thick leads than in areas with thin leads. This poor bonding should show up in lead pull surveys as an inordinate number of lead lifts and un-bonded leads for processes like L_1 which plate thicker leads on average (see Appendix A for a description of the lead pull test. Appendix A also provides a description of the three common failure modes observed in the lead pull test).

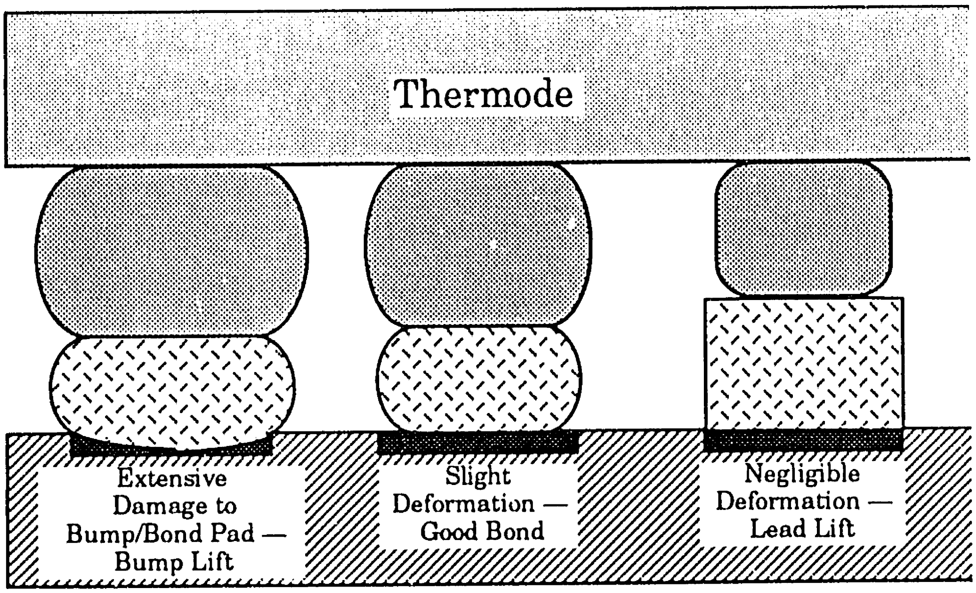
The second reason lead thickness variation might cause poor bonding is that, if the thermode and the die are coplanar but the inner leads are not, thick leads will bear more of the bond force than thin leads. The thermode will crush thick leads more than it crushes thin leads (see Figure 3.2). Thick leads should tend to bump-lift as their aluminum bond pads crack under excessive pressure or as extruded bump material cracks passivation. Having seen less pressure, thin leads adjacent to thick leads should tend to bond weakly and fail by lead lifting. Since L_1 allows more lead thickness variation than L_2 (standard deviation in lead thickness from Table 2.1 and Table 2.2), lots plated using L_1 should exhibit both more bump lifts and lead lifts than lots plated using L_2. Figure 3.2 shows the difference in crush between thick and thin leads (following page).

FIGURE 3.2 Pressure Difference Due to Lead Thickness Variation

(a) Prior to Application of Pressure



(b) Post-Application of Pressure



CHAPTER 3: ILB Yield Studies

3.1) LEAD PULL STUDY — GENERAL METHOD

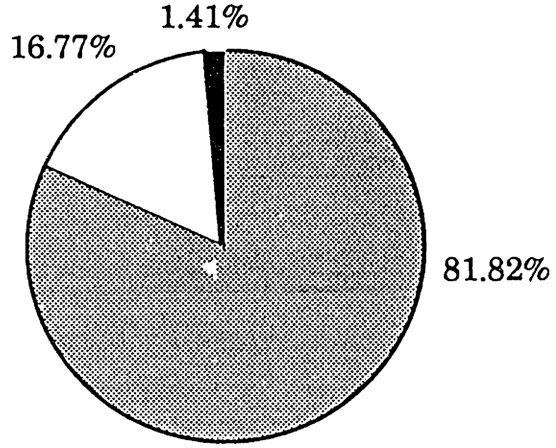
The source of all the information which follows in this chapter is the lead pull survey. All the data used were generated using the techniques outlined in Appendix A. Data was collected from engineering lots of TAB tape over a three-month period from June 1 through August 31, 1989. Overall, more than ten thousand data points were input by hand into the master-database. The procedure used in this investigation was to sort data along different parameters to explore trends in the data. Findings are divided into three categories: overall yield variation between lots plated using L_1 and L_2, differences in lead-to-lead yield within individual tape sites for one method vs. the other, and the improvement in the yields of both methods over time.

3.2) ILB YIELD VARIATION WITH LEVELING AGENT

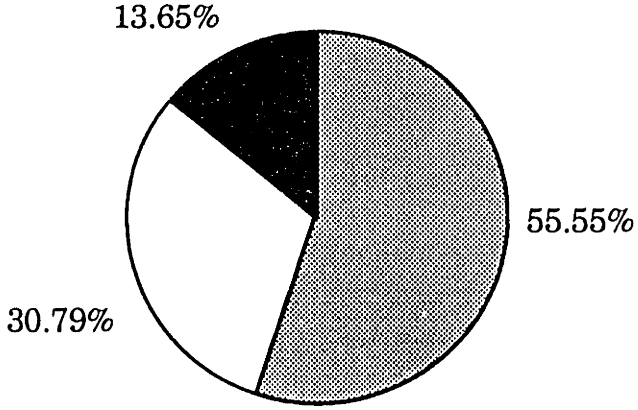
Preliminary analysis indicated that there is a difference in the frequency of the different failure modes described in Appendix A for lots plated using L_1 and lots plated using L_2. The ratios of lead breaks, lead lifts and bump lifts as a fraction of total lead pulls were calculated for both categories. Lead lifts are twice as frequent for L_1 as for L_2. Bump lifts are an order of magnitude more frequent. Figure 3.3(a) shows the breakdown for L_2. Figure 3.3(b) shows the same chart for L_1.

FIGURE 3.3: Failure Mode Frequency and Leveling Agent

(a) New Leveling Agent (L2)



(b) Old Leveling Agent (L1)

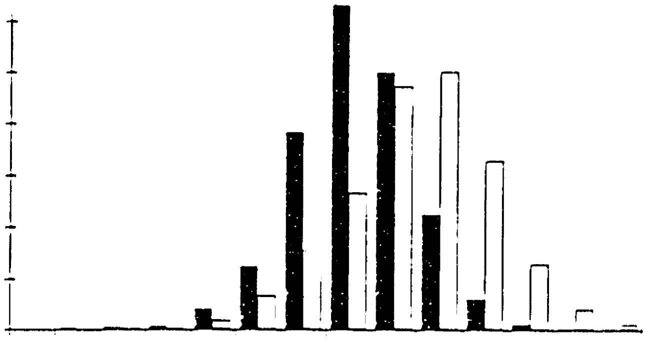


One trend which had not been anticipated was an increase in lead pull force for leads plated using L_2 to fail vs those of L_1. Explanations for the pull force shift are presented in §3.5. Figure 3.4 shows how the normal distributions shifted with respect to force for each failure mode[•].

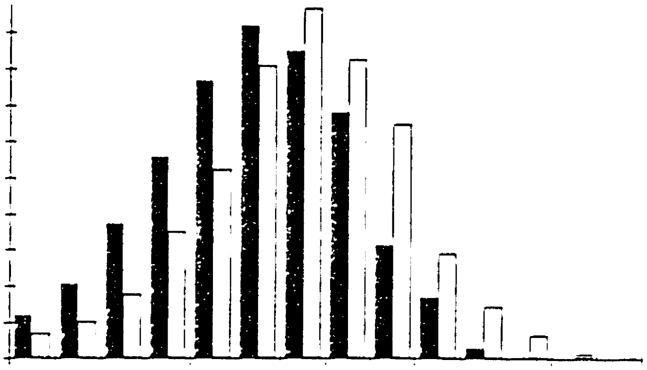
[•] Pull forces removed to protect proprietary information.

FIGURE 3.4: Pull Force Distribution vs Leveling Agent

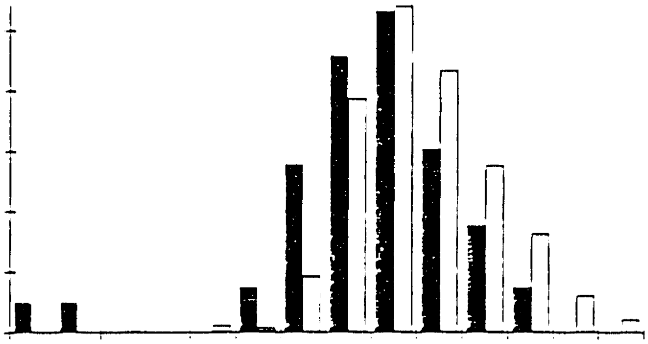
(a) Lead Break Distribution



(b) Lead Lift Distribution



(c) Bump Lift Distribution



□ L1 Tape Sites ■ L2 Tape Sites

To get a quantitative number for the difference in quality between L_1 tape sites and L_2 sites, a WALD chart was prepared comparing the two regions (see Appendix A for a description of WALD weighting). Figure 3.5 plots the cumulative WALD for both types of site from June 1 to August 31, 1989*. The line for L_1 is significantly higher than that for L_2. Bonding quality is therefore "worse" with L_1 than with L_2.

FIGURE 3.5: WALD Weighting vs. Leveling Agent

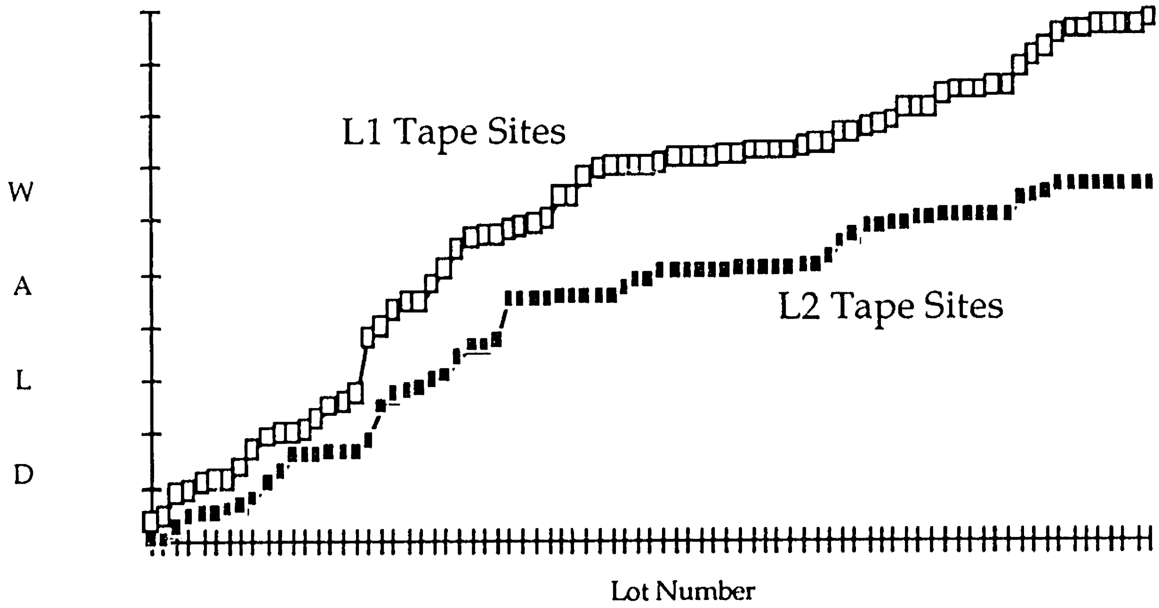


Figure 3.4 shows that the average pull force of lead lifts is higher for L_1 than for L_2. Therefore the WALD value should be inversely lower. The fact that it is higher is a reflection of the trends shown in Figure 3.3 that frequency of lead lifts is very much greater for L_1 than for L_2.

---

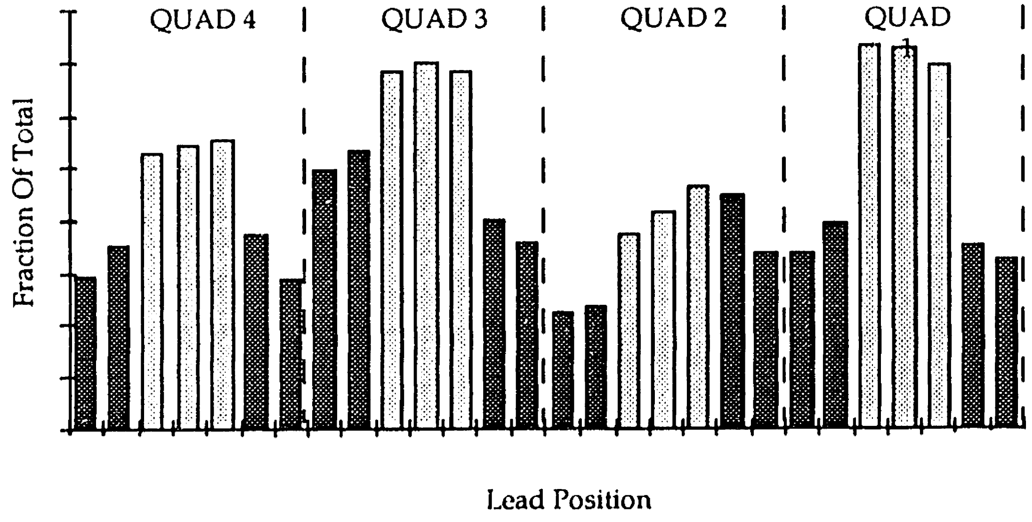
* The numbers have been removed from the axes to protect proprietary information.

3.3) ILB YIELD VARIATION vs. LEAD POSITION WITHIN TAPE SITE

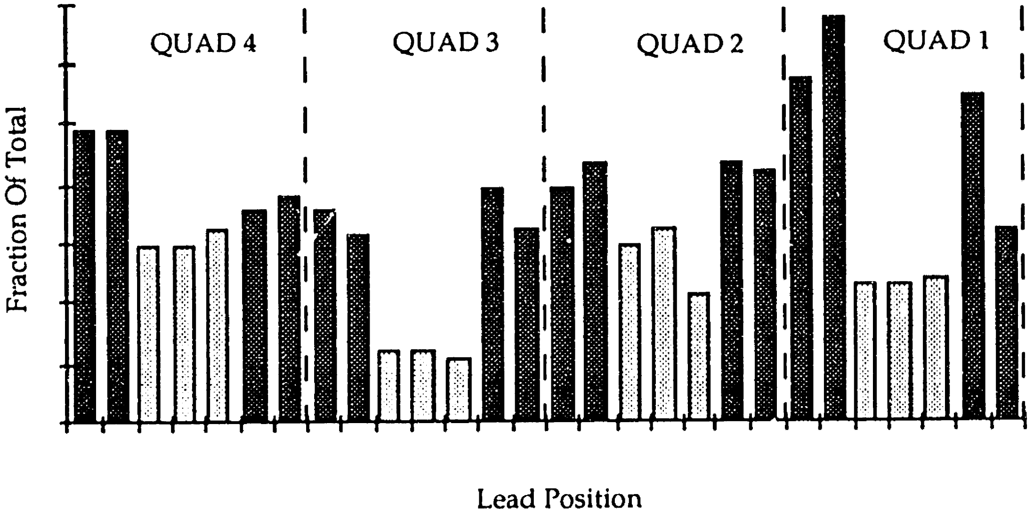
To check the role of intra-site lead thickness variation, lots with a relatively high degree of intra-site variation (i.e. lots plated using L_2) were analyzed for the frequency lead lifts and bump lifts as a function of the lead position within the tape site. For this investigation, the same lead pull data used above was batched according to the twenty-eight positions on the lead pull survey (see Appendix A). From the analyses performed in Chapter 2, it was known that when lead thickness variation is significant, leads tend to be taller in the corners of the quadrants than in the middles of the quadrants.

FIGURE 3.6: Failure Mode vs. Lead Position Within Individual Tape Site

(a) Lead Lift Distribution



(b) Bump Lift Distribution



As predicted using the model outlined in Figure 3.2, lead lifts are most frequent in the middles of the quadrants. Bump lifts appear to be concentrated in the corners of the tape sites. This trend corresponds well with the theory that thick leads in the corners transfer more force to the bond pads, cracking the pads and causing bump lifts, while thin leads see less force and lead lift.

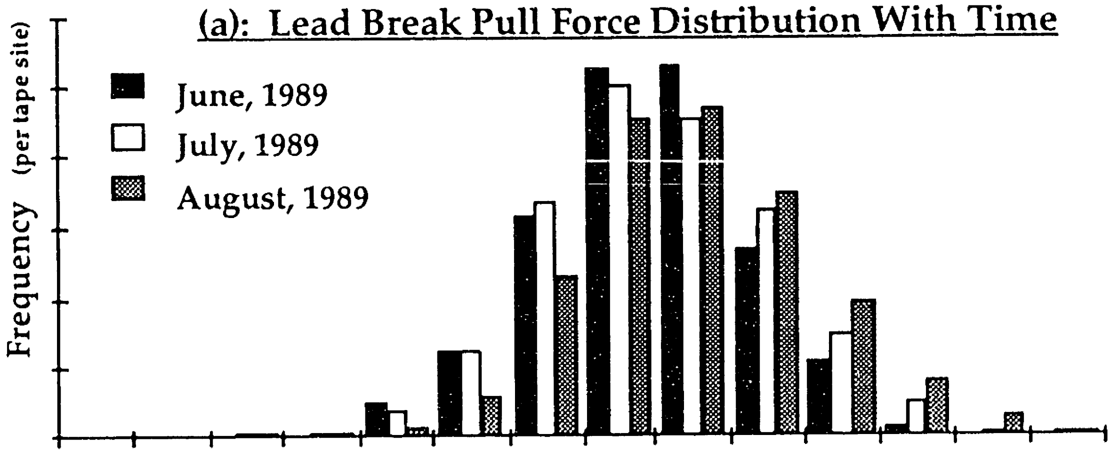
3.4) ILB YIELD CHANGE WITH TIME

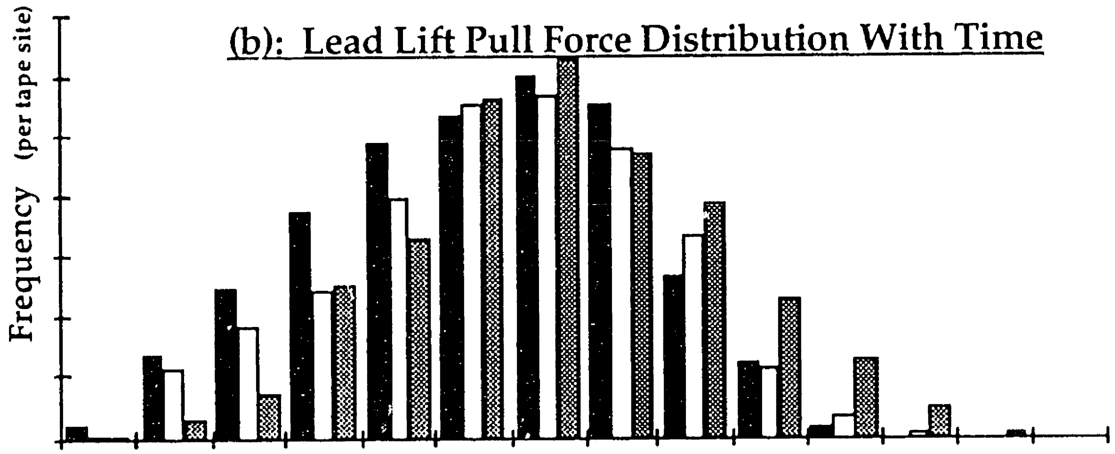
From lead pull data alone, it is impossible to differentiate between quality changes over time between the ILB assembly operation itself or incoming tape. Most likely, it is a combination of both: the tape quality improves <u>and</u> Motorola operators get better at optimizing parameters for TAB tape. It is possible to demonstrate trends in overall quality.

One indicator of quality is overall improvement in pull force distributions with time. Figure 3.7 shows how lead break, lead lift, and bump lift force distributions shifted over time (following page). Pull force distributions for lead breaks, lead lifts and bump lifts, respectively, shifted higher each month. There are three possible implications: (1) Motorola's bonding process improved; (2) quality improved overall; and, (3) TAB leads' tensile strength increased. Most likely, it was a combination of the three. These apparent improvements are discussed in §3.5.

To get a feeling for how the shift in pull force distributions impacts on product quality over time, another WALD chart was prepared using the weighting scheme outlined in Appendix A. Figure 3.8 plots cumulative WALD with total number of parts surveyed for the three-month period studied (on page 14). There are at least two distinctly different slopes to the data plotted there. About half-way through July, the quality markedly improved.

FIGURE 3.7: Improvement in Pull Force Distributions Over Time


(a): Lead Break Pull Force Distribution With Time


(b): Lead Lift Pull Force Distribution With Time

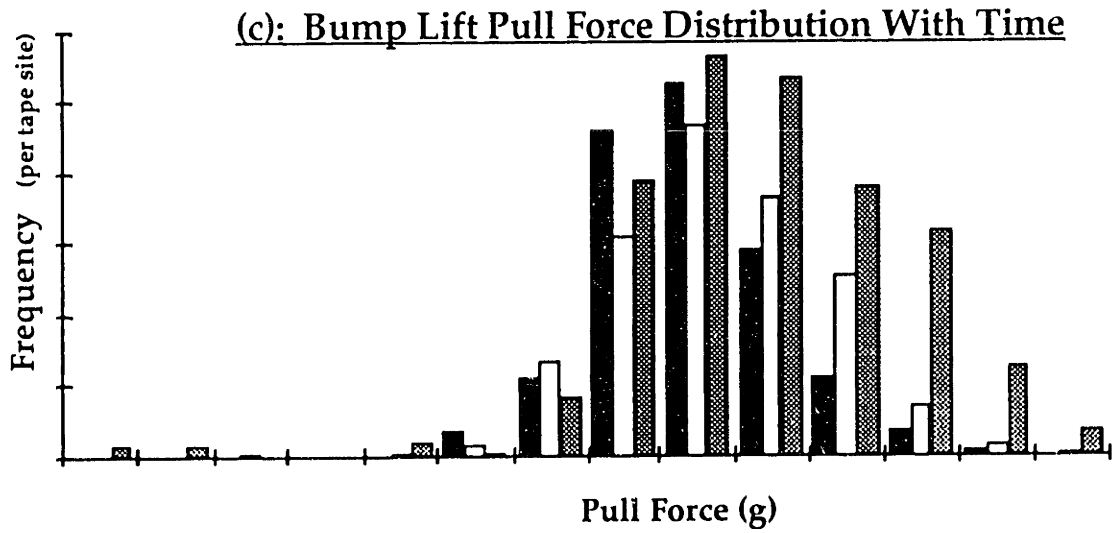
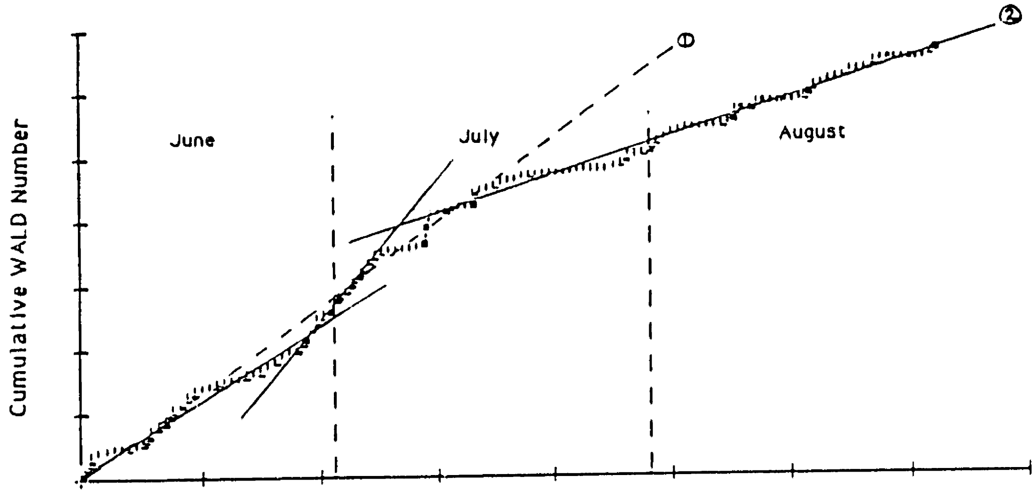

(c): Bump Lift Pull Force Distribution With Time

FIGURE 3.8: Total Cumulative WALD for Lead Lifts



3.5) SUMMARY

3.5.1) DISCUSSION

One unanticipated trend in the data is the correlation in Figure 3.4 of pull force with leveling agent — average pull force is higher for L_1 than for L_2. A potential explanation for this correlation is that L_1 sites, being thicker and cooler at the tin interface, are traditionally more difficult to bond than L_2 sites. Bonding parameters for L_1 sites are typically higher than bonding parameters for L_2 sites. The result is that they use more driving force for L_1 than for L_2 so that, overall, bond strength is higher for L_1. A better explanation is that since leads tend to be thicker with L_1, they are structurally stronger than leads plated using L_2. They will break at a higher pull force. With the lead break threshold higher, other failures are likely to occur before the ultimate tensile strength of the lead is exceeded. Therefore, more lead lifts and bump lifts should be seen with L_1 — which is case, as shown in Figure 3.3. In this situation, increased numbers of lead lifts and bump lifts might be an artifact of the lead pull test and not necessarily indicative of poor bonding. However, the WALD chart shown in Figure 3.4 indicates that bond quality as measured by lead lift force and frequency is, in fact, lower for L_1 than for L_2. The frequency of lead lifts using L_1 is so much higher than using L_2 that the higher occurrence of <u>low</u> lead lifts at the tails overshadows the overall higher pull force of L_1 with respect to L_2.

The high frequency of lead lifts and bump lifts of L_1 over L_2 corresponds well with predicted behavior based on Figure 3.1 and Figure 3.2. The lead and bump lift trends within individual tapes sites shown in Figure 3.6 particularly supports the hypothesis that local differences in bonding pressure caused by variation in lead thickness play a significant role in bonding yield.

Section 3.4 leaves demonstrates that tape quality improved over the course of the lead pull evaluation. Some of the improvement in bonding yield may be attributable to adjusting-up bonding parameters for L_1 tape sites over time — August showed an increase over July in the total frequency of bump lifts (Figure 3.7).

3.5.2) AREAS FOR FURTHER STUDY

One effort that still deserves attention is a characterization of bonding temperature as a function of lead thickness and thermode temperature. How does Figure 3.1 actually appear? Such an investigation should be tied to a quantitative characterization of the mechanisms involved in bonding. What steps are rate-limiting in the eutectic reaction describe in Chapter 1?

One reliability issue that should be studied is the proper WALD weighting associated with lead pull. Another issue that should be investigated is whether or not a WALD weighting scale should be assigned to bump lift failures and to lead breaks. The specification referred to in Appendix A lists binary failure criteria for both bump lifts and lead breaks as well as for lead lifts. There ought to be a way to weight these criteria quantitatively with pull force — just as there was for lead lifts. This additional information should provide Motorola with a better picture of their quality than just the lead lift information. As ILB quality improves, the overall frequency of lead lifts will drop. Lead lift WALD will cease to be a useful metric for process control. Process control will need to be based on a more typical failure — lead breaks.

Several critical variables have changed since the conclusion of the work that contributed to Chapter 2 and Chapter 3. For instance, Motorola has installed new inner lead bonding equipment. The types of data collected for Chapter 2 and Chapter 3 would be very useful in assessing the impact of these changes. WALD plots similar to the one presented Chapter 3 would be very useful.

3.5.3) CONCLUSIONS

(1) ILB yield loss correlates closely with leveling agent. This behavior correlates closely with observed inner lead thickness variation detailed in Chapter 2.

(2) ILB yield loss patterns correspond well with behavior predicted on the basis of the two models shown in Figure 3.1 and Figure 3.2. These initial indications should encourage further development of these two models.

(3) Overall product is steadily improving with time.

CHAPTER 3: ILB Yield Studies

1. G. H. Geiger and D. R. Poirier. Transport Phenomena in Metallurgy, 2nd Printing, 1980, §9.2.2, p. 278, Addison-Wesley Publishing Company

CHAPTER 4: DECISION MODELS

"Most human decision making, whether individual or organizational, is concerned with the discovery and selection of satisfactory alternatives; only in exceptional cases is it concerned with the discovery and selection of optimal alternatives. To optimize requires processes several orders of magnitude more complex than those required to satisfice."

James March and Herbert Simon
Organizations, 1958[1]

How do managers and engineers solve problems in the Hyperchip ramp-up environment? What role do pressure and of uncertainty play in framing their decisions? This chapter summarizes some of the models presented in management literature which have addressed these and similar questions.

March and Simon recognize the impossibility of eliminating uncertainty from the decision making process. Their insight has not stopped management scientists from advancing a multitude of optimizing aids. Inventory scheduling, linear programming, economic cost modeling, multivariate utility analysis, capital gains theories and many other techniques all promise to help decision makers make the most appropriate choices. Some tools have been more successful than others. Japan has become a dominant force in global economics largely because decision rules developed by its industries (which are often based on improving rather than on optimizing) have proven more responsive to the market than those used by their Western counterparts. Yet the abundance of decision aids does not alter the fundamental truth of March and Simon's words: it is prohibitively difficult to always make the best decisions.

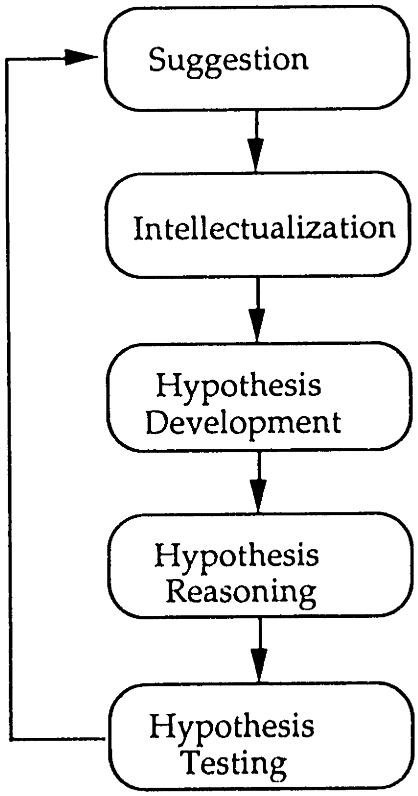
Optimizing models only work under limiting assumptions. The most basic of these is certainty. March and Simon define certainty as "quantifiable risk" - a priori knowledge of all possible results and their likelihood of occurring. An estimation of expected value dictates the optimal decision.

Decision makers usually do not have the luxury of certainty. More often, they operate in situations in which there is prior knowledge neither of all possible solutions nor of how to weight them. Several authors have pointed out that under conditions of uncertainty, there is no single best solution[2]. Moreover, decision makers often lack the resources to develop certainty before external pressures force them to act on a decision. Many authors have dealt extensively with the structure of decision making under both uncertainty and pressure. There are two primary categories of uncertainty: situations in which not enough information exists to establish certainty and situations in which so much information exists that it is difficult to determine which information is valid and which is not. Several authors have suggested screening techniques decision makers use to deal with large amounts of information. Some of these are discussed below.

CHAPTER 4: DECISION MODELS

Situations of uncertainty preclude optimization, but not systematic decision making[3]. Early decision making models loosely imitated the Scientific Method. John Dewey's "Five Phases of Reflective Thought" (Figure 4.1) developed in 1910 is a typical example[4]. Like the Scientific Method, the model is an incomplete representation of the problem solving process in that it is strictly diagnostic rather than prescriptive. Given enough time, it can detail the nature of a problem, but will not in itself suggest a solution.

Figure 4.1: Five Phases of Reflective Thought

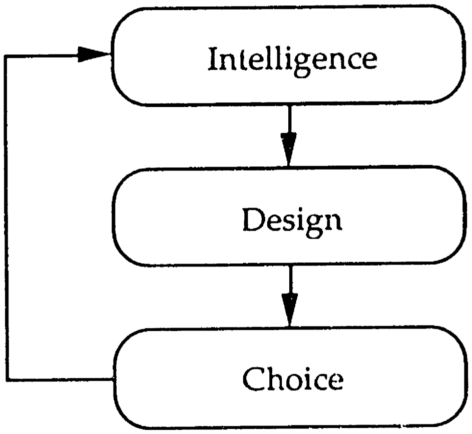


To describe more completely how people solve problems, Herbert

CHAPTER 4: DECISION MODELS

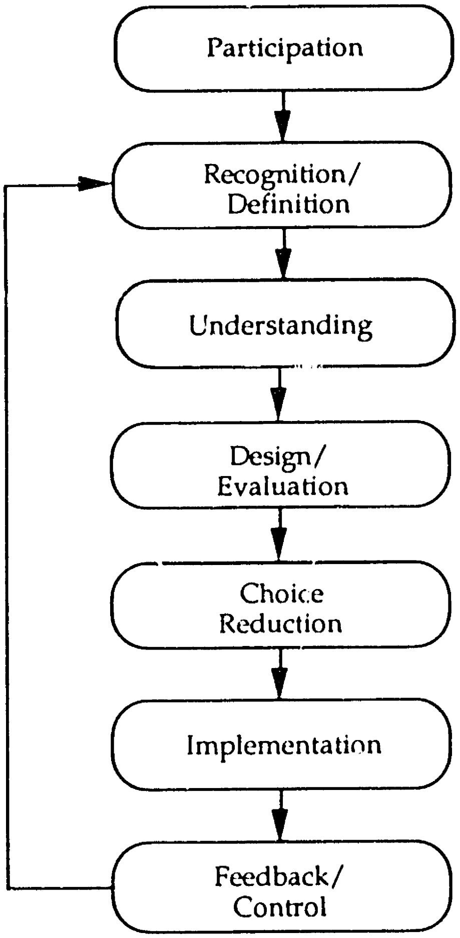
Simon proposed his now classic three-tiered decision making framework: intelligence, design, and, choice[5]. "Intelligence" is becoming aware that a problem exists - in Mintzberg's words, "finding occasions for making decisions." "Design" includes locating and/or generating a range of solutions and subsequently evaluating them. "Choice" is selecting and implementing a single alternative[6]. Figure 4.2 shows Simon's framework:

Figure 4.2: Simon's Decision-Making Trichotomy



Most modern problem solving models build upon Simon's basic trichotomy. Nearly every model explicitly includes problem recognition, search or generation of alternatives, and the evaluation and selection of the most appropriate alternative. Most, like Soelberg's in Figure 4.3, also include feedback mechanisms at various critical points. The structure of the rest of this chapter roughly follows the three elements of problem solving: recognition of a problem, search for solutions and selection of a single alternative.

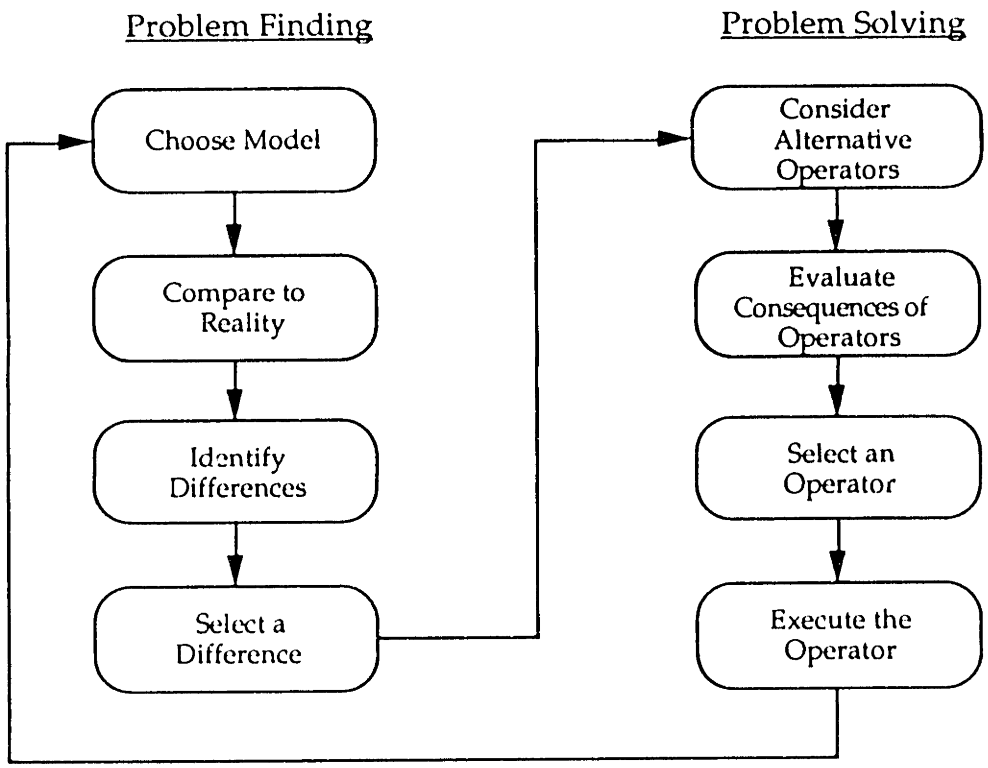
Figure 4.3: Soelberg's Unprogrammed Decision-Making Model



4.1) PROBLEM RECOGNITION

To say something is "bad", one must answer the question "bad in comparison to what?" William Pounds defines "problem" as the difference between existence and a desired situation. Perception of these differences (Problem Finding) is the first step in Pounds' problem solving process (Figure 4.4):

Figure 4.4: Pounds' Problem Finding Model



4.1.1) Pounds' Problem Finding Templates

Problem Finding requires that the decision maker must: 1) be aware of the existing situation; 2) be able to envision of a better situation; and, 3) be able to judge the difference between the two (solving the problem, or reducing the difference, Pounds leaves as an exercise to the reader)[7]. For Pounds, the most important of these three elements is the second. He assumes that the manager will know what is the status quo and will be able to judge the differences between it and a more desirable condition. He thereby limits the decision maker's initial level of uncertainty. Pounds proposes four templates against which managers define problems[8]:

Historical. Historical templates assume that the recent past is the best estimator of future performance. When performance drops below the trend predicted by past events, the manager recognizes it as a "problem".

Planning. A plan is a projection into the future of appropriate managerial activity. If a manager misses milestones, fails to complete tasks by assigned deadlines, etc., it creates a "problem". Pounds feels planning templates are less important than the Historical Model since, in the company he studied, managers tended to operate well within their plans. As long as managers have little difficulty performing as well as required by a plan, they are free to define problems using some other criteria.

"Other People's" and Extra-Organizational. Customers, upper level managers, competitors and regulations all may directly impose external models on a manager's activity. If performance falls below that required by external models, third parties may define a "problem" for the manager.

Scientific. Managers and scientists both use models to predict behavior. Pounds brings up Scientific Models to contrast how managers react to deviations from their models with how scientists and engineers react. When reality deviates from performance predicted by a managerial template, the manager defines a "problem" and attempts to alter reality to match the model. When reality deviates from behavior predicted in a Scientific Model, the researcher detects a "problem" and attempts to modify the model to more closely approximate reality.

Lyles and Mitroff agree with Pounds that the most important step in solving problems under uncertainty is framing the problem. However, they feel that most managers become aware of a problem through informal channels prior to being alerted to it by Pounds' templates[9]. In his own treatment of problem solving, Soelberg outlines four templates similar to Pounds': 1) The discovery of a barrier to progress toward an objective; 2) A request from another person or company; 3) A performance indicator that drops below a target level, and; 4) The perception of a pattern that was previously classified as a "problem". One distinction that Soelberg makes is that in order to perceive any of these situations as problems, the decision maker must first be a participant in the process (see Figure 4.3) — a manager must have a sense of ownership[10].

4.1.2) Stimuli for Decision-Making Activity

In an early work, Chapman, etal. suggested that decision makers detect a problem when certain environmental stresses exceed some critical threshold[11]. They identified two primary stresses: discomfort stress and failure stress. An organization feels discomfort stress when a process meets minimum requirements, but does not work as well as it could. An organization experiences failure stress when performance can not meet minimum levels of acceptance at all. Chapman felt that managers most often noticed problems based on their level of discomfort, as opposed to their fear of failure*. The chief role of failure stress in his model was to screen potential solutions for minimum levels of acceptance. He generalized this situation to most business environments[12].

Different authors have felt that stimuli other than discomfort are more likely to motivate problem solving activities. Five years after Chapman published his work, Cangelosi & Dill defined a third stress: disjunctive stress. Disjunctive stress arises from tensions between groups and individuals due to differing expectations. They felt most "problems" could be defined as the product of tension between groups[13].

Even later, other authors concluded that the threat of failure induces problem solving activity more often than any other stress[14]. Mintzberg classifies decision processes as opportunities, crises and problems

---

* Chapman's research centered on a very homogeneous organization with a single common goal: an Air Force early warning station with the fundamental goal of defending the country from potential air strikes. This study group was one in which everything worked, but not as well as it could.

depending on the stress that prompts them**. Mintzberg agrees with Jackson, etal., that threats and problems spur decisions more often than do opportunities. He concludes that even in fairly stable environments, the majority of decision processes center around either "serious" problems or crises. Mintzberg also notes that only in rare cases is problem solving undertaken on the basis of one stimulus (natural disasters, deaths, takeover attempts, etc.). Usually an accumulation of multiple small stimuli punctuated by one intensive stimulus motivates decision making.

4.1.3) Problem Definition: Bounding the Decision Space

After problem noticing, the second step in "Recognition" is definition. Pounds assumes that managers define problems automatically as part of contrasting reality with a template. Soelberg believes an important part of problem definition is establishing goal dimensions along which to evaluate alternatives. Soelberg and Pounds think decision makers subconsciously specify these dimensions as a natural part of contrasting a situation with a model[15].

Mintzberg suggests that soon after decision makers detect a problem

---

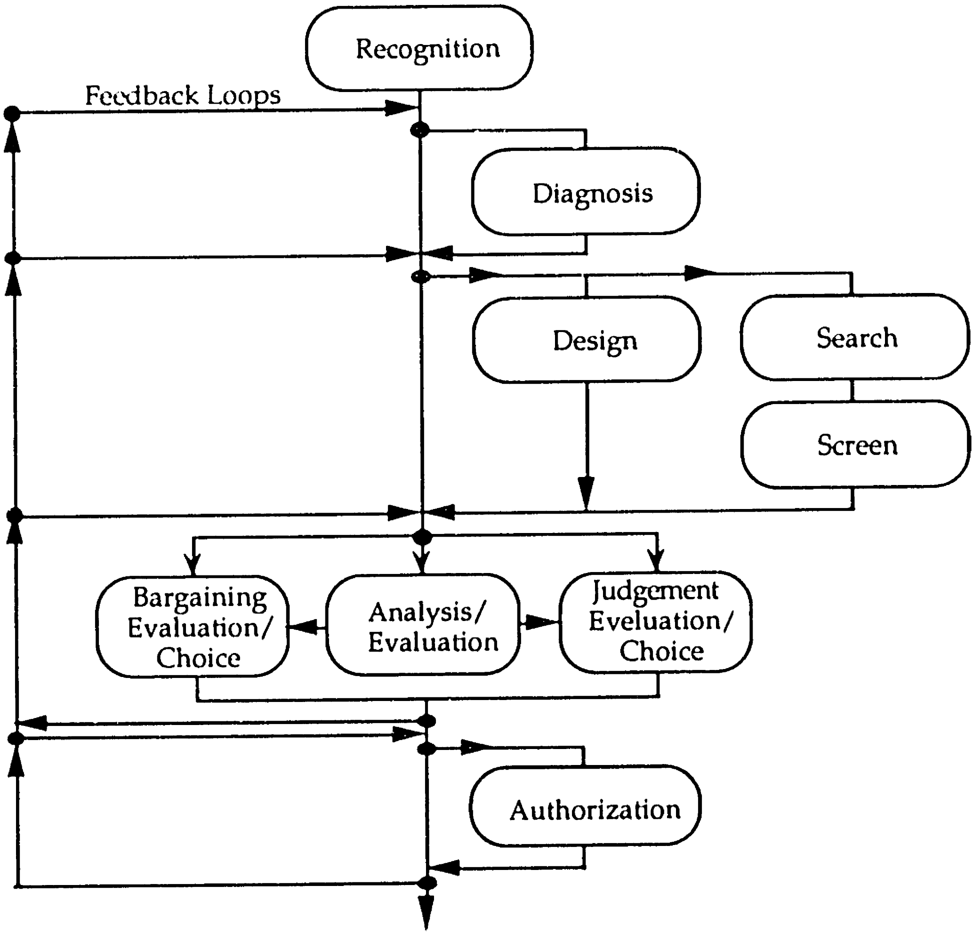
** An opportunity decision is motivated by desire to improve an already secure situation. A crisis decision is motivated by "intense pressure demanding immediate action." "Problems" fall somewhere in between. The nature of a stimulus can change from opportunity to problem to crisis as decision-making activity is delayed. Conversely, the severity of a crisis may be minimized by invoking temporary solutions.

they use intuitive reasoning to define first-order cause-and-effect relationships among obvious variables. In Figure 4.5, the "Diagnosis" step immediately follows Recognition. Mintzberg's Diagnosis routine consists primarily of decisions about how to conduct the problem solving process — establishing a schedule, determining which and how many resources to commit to the solution, etc.. The decision path divides just before Diagnosis, indicating that under some circumstances, this step is bypassed. One situation in which Diagnosis might be skipped would be one where a problem encountered is very similar to previous ones so variables are already well understood and an appropriate response is already established. Another scenario would be one in which very tight time limitations force the decision maker to streamline the process at every opportunity[16].

4.2) SEARCH FOR POTENTIAL SOLUTIONS

The decision maker identifies possible solutions to a problem during the "Search" phase. Soelberg's Search begins locally as exploration (passive scan for information) and spreads outward as investigation (focused search for specific information) as the problem remains unresolved[17]. Mintzberg thinks that decision makers begin searching for solutions using the easy techniques and progress to more resource-intensive activities as the easy ones fail to uncover viable alternatives. He lists four hierarchical levels of the search phase: 1) Memory Search, 2) Passive Search, 3) Trap Search, and, 4) Active Search. Memory search is a scan of organizational memory, including both personal memory and stored knowledge. If the memory scan fails to uncover a solution, decision makers progress to the next search activity. Passive search is waiting for unsolicited solutions to present themselves. Trap search involves the activation of "search generators" — managers send out feelers outside their organization and wait until some external group with a stake in solving the problem proposes an acceptable solution. In active search, decision makers directly seek alternatives. They experiment to develop new insights, set up prototype lines, etc.[18].

Figure 4.5: Mintzberg's Iterative Problem Solving Model



CHAPTER 4: DECISION MODELS

The goal of Mintzberg's Search is to discover ready-made solutions. As the search uncovers possible solutions, the decision maker screens them for minimum levels of acceptability. March & Simon call this screening activity "satisficing". During screening, the decision maker builds an active roster of satisfactory alternatives[19]. The search progresses from memory search to active search as first one method then the next fails to unearth a viable solution. If the search reveals no satisfactory alternatives, the problem solver must design a custom solution or modify an existing one to meet the specific problem[20].

Neither Mintzberg nor Soelberg believes the search routine is truly sequential. Both think it is artificial to separate individual steps in the process and to imply that one is carried out after the other. For them, it is an iterative process that takes many cycles (Mintzberg's model - Figure 4.5 - is full of feedback loops). They feel that: (1) search begins as soon as a problem is identified; (2) alternatives arise throughout the search; (3) automatic screening for an "active roster" begins as soon as alternatives are generated; and, (4) evaluation takes place immediately on alternatives in the active roster.

If the search fails to produce any minimally acceptable alternatives, the decision maker faces three choices: (1) quit; (2) change the definition of "minimally acceptable"; or, (3) design a tailored solution. Quitting is only an option for opportunity decisions or for minor problems, not for crises. Designing a tailored solution is usually resource-intensive, in terms of time, money, equipment and human capital. It is so costly that Mintzberg suggests only one tailored solution is ever made for a particular problem (that is why, in Figure 4.5, the design option is not followed by a screening step, as is the search option). The easiest way out is to change the definition of minimally acceptable. While this action may seem to be a "cop-out" and is clearly not vogue in an age of total process improvement, it often represents the only realistic choice available to a decision maker confronted by a complex problem and narrow time constraints.

After defining a problem, the decision maker determines if enough is known about the situation to initiate problem solving activity. Roger Bohn suggests that the understanding of any one variable may be roughly quantized. His "states of knowledge" range on a scale of zero to eight from total ignorance to complete characterization of a problem and its solution[21]. Assuming the decision maker is comfortable with the amount of knowledge about a problem, the next step is to choose a specific solution.

4.3) SELECTION OF AN ALTERNATIVE

4.3.1) EVALUATING ALTERNATIVES

Ideally, the selection criteria for solutions maximize the chances of "success". One of the difficulties faced by decision makers is how to measure success. March & Simon's first requirement for optimization is a set criteria by which all alternatives may be evaluated; there must be a common yardstick to judge which solution is "optimal"[22]. Because it is difficult to objectively measure success, decision makers use a number of shortcuts to help them make choices. One is satisficing in the search phase. Another is making decisions by proxy. A manager may select an alternative based on imitation, tradition or the personality of the alternative's sponsor rather than on the merits of the alternative itself. Or, the manager may substitute "soft" issues for quantitative goal dimensions[23]. Problem solvers often invest as much faith in opinions as in facts and select alternatives on the basis of emotions, politics or bias. Kiesler & Sproull outline ten proxies used by managers to weight data[24].

Cangelosi and Dill suggest that existing goal dimensions are often arbitrary, providing little true guidance. Many of them are "temporary" guidelines which have taken on permanence due to organizational resistance to change[25]. Decision makers rarely establish explicit goal dimensions for specific problems they address. Then they rarely apply yardsticks equally to alternative solutions. They usually evaluate alternatives independently against non-comparable goals. They then select a favorite alternative based on not more than two or three primary dimensions. They make this selection implicitly — often without being fully aware of it — before the formal selection phase begins and often before they are willing to discontinue the search. They spend the rest of the search seeking reasons to justify the favorite over all other potential solutions. During the selection phase, they use secondary dimensions to justify the favorite. Soelberg calls this rationalization step the confirmation process[26].

4.3.2) CONFIRMING THE SELECTION

Soelberg studied decision making as it pertained to business school graduates selecting a job. This situation was one in which the decision makers had relatively little information on which to base a major decision. The uncertainty they had about making their decision was the source of a great deal of anxiety for them. Therefore, before they were willing to announce a decision (or even admit to themselves that they had made a decision), they wanted some sort of reassurance that their selection was the best choice.

The confirmation process resolves such residual uncertainties. The decision maker tailors secondary goals and decision rules to best demonstrate the pareto dominance of the implicit choice over all the other alternatives. An appropriate decision rule is constructed ex post facto to explain this superiority. Soelberg believes the decision maker never voluntary stops the search process before at least one lesser alternative is found to serve as a straw man for confirmation. Conversely, if the search has stopped before time and resources run out, the decision maker will have already selected a favorite and a straw man. The formal selection phase serves primarily as ratification of a pre-selected alternative. When time and resource constraints force disclosure of the favorite before a suitable defense has been formulated, the decision maker undergoes a great deal of stress, having to support a choice whose dominance can not easily be demonstrated[27].

Many reasons for the confirmation process are external. Mintzberg discusses three formal selection routines: judgement, bargaining, and authorization. A supporting routine, analysis, parallels Soelberg's confirmation process (Figure 4.5). Its purpose is to build supporting evidence for a favorite alternative. Judgement is a personal evaluation from a single decision maker not subject to external review. In this situation, the decision maker's use for the confirmation process is limited to personal reassurance that the favorite alternative is in fact the best. Bargaining and authorization routines imply external certification of a decision. In these cases, the decision maker must defend a favorite solution against competing solutions (bargaining) or to a presumably skeptical senior manager (authorization)[28]. The confirmation process serves as the basis of the defense.

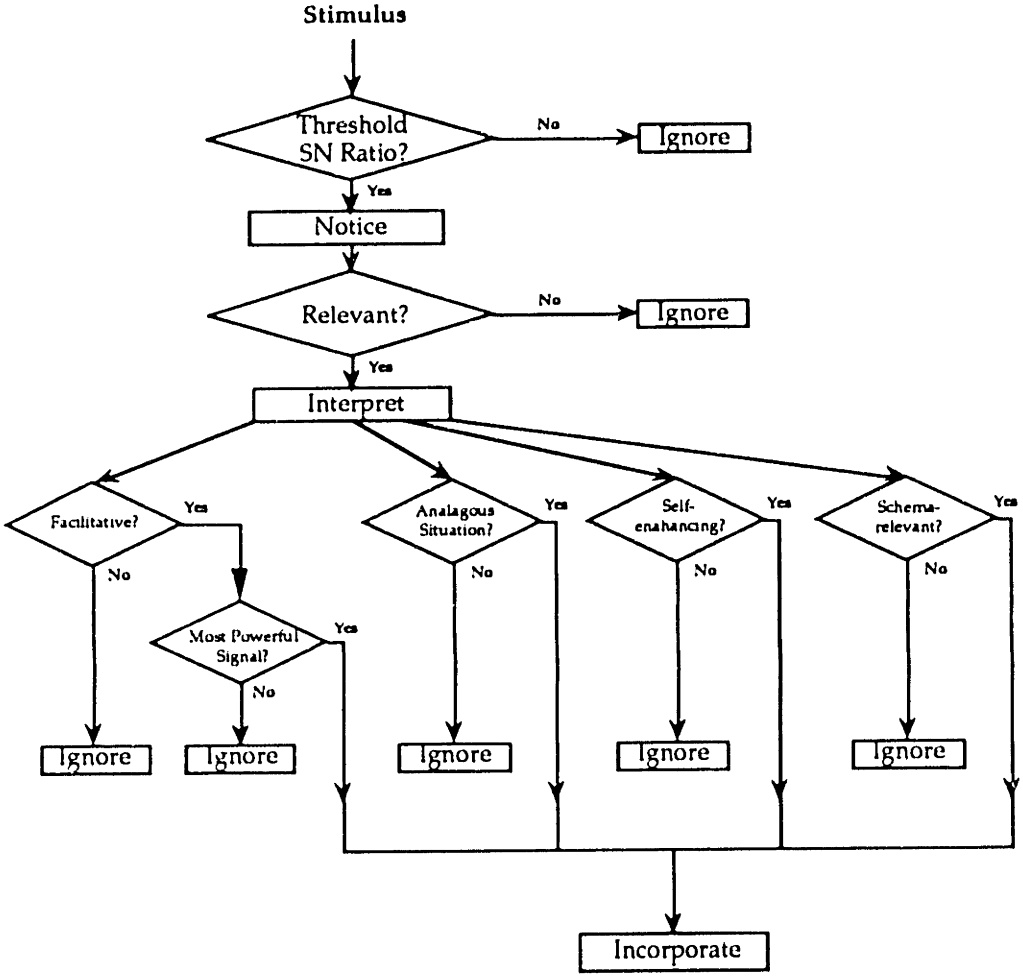
4.4) INTERPRETATION OF DATA

Many authors assume the critical factor in uncertainty is the quantity of information available to managers. However, in his definition "of states of knowledge", Bohn recognizes that the quality of knowledge is also very important[29]. Most problems are situations in which managers confront many sorts of information from many sources. The decision maker's first task is to identify which information is relevant in defining and solving the problem. Thoroughly screening all available information by objective metrics is often unrealistically time-consuming and costly.

In response to time and resource constraints, decision makers subconsciously use subjective shortcuts in assessing the importance of data. Because these shortcuts are irrational means of valuing data, some authors view them having primarily a negative effect on the decision process — they are traps to be avoided rather than tools to be used. Kiesler & Sproull identify many of these shortcuts. They call them "interpretive pitfalls" because they potentially lead the decision maker to incorrect assumptions throughout the problem solving process. Figure 4.6 shows five ways in which managers may misinterpret data. According to Kiesler and Sproull, Managers may <u>ignore important information</u>, <u>include unimportant information</u>, and <u>mis-interpret information</u> which they ultimately incorporate. They make these errors because they use the shortcuts shown in Figure 4.6[30]:

> <u>Facilitative Information</u>. People tend to give more credence to information which directly aids in defending a position. We downplay the significance of information which is not "facilitative". Lyles and Mitroff similarly found that managers tend to pay more attention to data which agreed with positions to which they are committed.

Figure 4.6: Kiesler & Sproull's Decision Model —"Interpretive Pitfalls"



Powerful Stimulus. If there is more than one stimulus, people tend to overestimate the significance of the largest. We tend to ignore all but the most powerful signal of a range of inputs.

Analogous Information. Another way in which we subconsciously define "important" information is by checking to see if similar situations have come up in the past. If decision makers can not find "analogous" information, Kiesler & Sproull believe they ignore it.

Self-Enhancing Information. The credibility we give information also tends to be directly related to the number of times we encounter it. Managers are more willing to accept as true information they hear several times from different sources - i.e. "self-enhancing" - than information they hear only once.

Schema-Relevant. We also tend to give more weight to data which fits nicely into our view of the world than that which does not. managers tend to overestimate the significance of information which is "schema-relevant".

One other criterion not listed on Figure 4.6 is the source of the information. Managers tend to weight heavily information given by "credible" or powerful sources[31]. Lyles and Mitroff found source credibility to be the most influential variable in determining the relevance of information. A critical factor in determining the credibility of the source is the quality, amount and format of empirical data that person has gathered to defend his or her position[32].

However, Kiesler and Sproull's treatise on decision-making shortcuts is negatively biased. Such shortcuts are not necessarily bad. Under twin stresses of uncertainty and time pressure, they are often the most reliable means a manager has to separate useful data from noise.

Chapter 5 present several examples which illustrate this point.

One of Bohn's assumptions is that as their understanding grows, managers become less susceptible to incorrectly using Kiesler and Sproull's shortcuts[33]. More knowledge permits decision makers to take specific actions with a higher degree of confidence in their solutions. Learning is an important aspect of reducing uncertainty. However, many of the same elements which introduce uncertainty into the decision making environment also hinder learning.

4.5) LEARNING IN THE PRODUCTION ENVIRONMENT

4.5.1) MOTIVATIONS FOR LEARNING

The stage of knowledge is a critical factor in every phase of the problem solving process. To recognize a problem, a manager must have a minimum level of awareness about a subject. To initiate an effective search, it is helpful to know where to begin to look. To weigh alternatives, the problem solver should be familiar with important goal dimensions.

In 1959, Chapman, etal. observed the learning process in the stable environment of a military air base. They view learning as a natural response to discomfort and failure stress and see problem solving as an off-shoot of learning. They note that learning tends to occur in discreet steps as tensions exceed threshold values. In their highly structured experiments, learning evidences itself as functional segregation, procedural shortcuts, increased selectivity of responses and implementation of redundant solutions at critical interfaces. They note that learning tends to "just happen". In their words, it "often evidenced itself abruptly and was not signaled by prior events." Their most important insight into <u>how</u> to learn is that organization needs to set goals to focus learning efforts[34].

In contrast to Chapman's study, Hirschman and Lindblom find that learning stems mostly from continual reaction to timely problems rather than from any programmed scheme. They write: "Learning occurs in response to immediate and obvious problems, imbalances and difficulties much more than it does as a derivative of some ex ante plan, theory, or ideology[35]." The source of the difference between Chapman's observations and Hirschman's is the environment they studied — Chapman explored a rigidly stable military establishment while Hirschman studied more fickle, rapidly developing organizations. The Hyperchip ramp-up more closely resembles Hirschman's environment than anything Chapman studied. The importance of the ramp-up environment is discussed below.

Mintzberg notes that organizational learning is based largely on verbal data. People spend much more time getting briefed than reading briefs. Communicating results to interested parties is a critical part of Mintzberg's model[36]. Soelberg notes that it is difficult to distinguish when learning is undertaken to aid in the search for a solution or in the definition of a decision rule from when it is undertaken to gather supporting evidence as part of a confirmation process. Kiesler & Sproull highlight that fact that learning may occur and yet not evidence itself, a la Chapman, as positive improvements in an organization. It may, in fact, not evidence itself as learning at all, but rather as a new problem if data is severely misinterpreted[37].

The goals of a ramp-up are dramatically different than those of a stable process. Bohn describes it as the difference between static and dynamic efficiency[38,39,40]. In a static environment such as a manufacturing operation for a mature product, typical goals are lower costs, stable processes, and optimization. Learning is motivated primarily by discomfort stress and is backed by extensive familiarity with critical variables. In a dynamic environment such as a volume ramp-up on a complex product and process, typical goals are fast development, rapid improvements in quality, introduction of process control, etc.. Learning is motivated primarily by failure stress - "problems" are all "crises" whose solutions are temporary firefighting fixes. The only issues addressed are current problems. The immediate needs of solving one problem draw resources away from resolving other current problems[41]. Learning is required and complicated by the fact that there is typically little initial fundamental product or process knowledge. Kiesler & Sproull's ten "propositions" can be particularly deadly in a dynamic environment. Thus, for Bohn, the most important management issue in a ramp-up is the management of learning[42].

4.5.2) BARRIERS TO LEARNING

The immediate needs of the production environment often pre-empt learning activities. Bohn suggests that in a fast ramp-up, firefighting creates a persistent crisis management mentality. Learning activities become dedicated toward meeting the next development milestone rather than focusing on long range issues of decreasing uncertainty. Tyre suggests there are indirect and direct costs to concentrating on problem solving or learning activities in a typical manufacturing environment. Indirect costs include be tradeoffs between engineering effort devoted to learning and problem solving and engineering support of process and quality control. This cost becomes particularly significant in situations where production runs through several shifts while technical support is available only during the day shift. Direct costs include lost machine and production time while factory equipment is used in experiments. Production managers also become reluctant to temporarily shut down even a low-yielding process for fear of losing the recipe[43,44].

Organizational constraints also limit learning. A significant cost is the risk of disruption of a stable working environment. Workers with limited attention and time "can attend to only some aspects of the environment." They are reluctant to invest the energy necessary to incorporate new data or to view old data in a new light. Learning and problem solving involves a higher level of effort than does "maintaining".

Therefore, operators and sustaining engineers tend to resist changes which do not visibly lead to increased stability and the entrenching of a routine[45].

A second significant organizational impediment to learning is the "Not Invented Here" syndrome (NIH). In a company made up of groups having differing priorities, groups and individuals inevitably tend to become territorial. A solution implemented by one group is not necessarily adopted by other groups strictly on the basis that the solution did not originate within their organizations. Local organizations therefore tend to "reinvent the wheel" rather than "beg, borrow or steal" ideas from external sources. NIH can spell disaster for companies with short life-cycle production methods or products. With their premium on speed, these companies cannot afford the delays and risk involved in inventing internal solutions for every contingency.

4.6) SUMMARY

Companies involved in ramp ups must quickly recognize and solve problems while constantly seeking to reduce uncertainty. Inevitably, as noise and uncertainty decrease, more problems become apparent which were previously unnoticed. These companies must therefore also be capable of weighting and prioritizing their problem solving activity as well as balancing short term firefighting activities with longer-term learning activities. Managers must constantly be aware of the tendency of problem solvers to complicate the decision making process by delaying their decisions until confirmation candidates can be evaluated. Similarly, they must walk a fine line between developing customized, highly optimized solutions in which they have a high degree of confidence and simply reinventing the wheel.

The following chapter describes the Hyperchip decision making environment. It outlines some of the pressures and uncertainties experienced by managers in this ramp-up and illustrates them with short anecdotes. The final chapter discusses possible ways to minimize uncertainty without sacrificing along short-term performance dimensions. It suggests new insights into the models proposed in Chapter 4 and proposes several new management priorities which are a direct implication of these new insights.

CHAPTER 4: DECISION MODELS

1. J. March and H. Simon. Organizations, John Wiley & Sons Pub., 1958.

2. M. Lyles and I. Mitroff. "Organizational Problem Formulation: An Empirical Study," *Administrative Science Quarterly*, p. 102-119, March, 1980.

3. P. Soelberg. "Unprogrammed Decision Making," *Industrial Management Review (renamed Sloan Mgmt Review)*, p. 19-29, Spring, 1967.

4. H. Mintzberg, R. Duru and A. Théorêt. "The Structure of 'Unstructured' Decision Processes," p. 246-275. *Administrative Science Quarterly*, Vol. 21, 1976.

5. H. Mintzberg, R. Duru and A. Théorêt. "The Structure of 'Unstructured' Decision Processes," p. 246-275. *Administrative Science Quarterly*, Vol. 21, 1976.

6. P. Soelberg. "Unprogrammed Decision Making," *Industrial Management Review (renamed Sloan Mgmt Review)*, p. 19-29, Spring, 1967.

7. W. Pounds. "The Process of Problem Finding," *Industrial Management Review (renamed Sloan Management Review)*, p. 1-19, Fall, 1969.

8. W. Pounds. "The Process of Problem Finding," *Industrial Management Review (renamed Sloan Management Review)*, p. 1-19, Fall, 1969.

9. M. Lyles and I. Mitroff. "Organizational Problem Formulation: An Empirical Study," *Administrative Science Quarterly*, p. 102-119, March, 1980.

10. P. Soelberg. "Unprogrammed Decision Making," *Industrial Management Review (renamed Sloan Mgmt Review)*, p. 19-29, Spring, 1967.

11. R. Chapman, J. Kennedy, A. Newell, W. Biel. "The Systems Research Laboratory's Air Defense Experiments," *Management Science*, p. 250-269, Vol. 5, 1959.

12. R. Chapman, J. Kennedy, A. Newell, W. Biel. "The Systems Research Laboratory's Air Defense Experiments," *Management Science*, p. 250-269, Vol. 5, 1959.

CHAPTER 4: DECISION MODELS

13. V. Cangelosi and W. Dill. "Organizational Learning: Observations Toward a Theory," *Administrative Science Quarterly*, p. 175-203, Vol 10, Sept., 1965.

14. S. Jackson and J. Dutton. "Discerning Threats and Opportunities," *Administrative Science Quarterly*, p. 270, Vol. 33, Sept., 1988.

15. P. Soelberg. "Unprogrammed Decision Making," *Industrial Management Review (renamed Sloan Mgmt Review)*, p. 19-29, Spring, 1967.

16. H. Mintzberg, R. Duru and A. Théorêt. "The Structure of 'Unstructured' Decision Processes," p. 246-275. *Administrative Science Quarterly*, Vol. 21, 1976.

17. P. Soelberg. "Unprogrammed Decision Making," *Industrial Management Review (renamed Sloan Mgmt Review)*, p. 19-29, Spring, 1967.

18. H. Mintzberg, R. Duru and A. Théorêt. "The Structure of 'Unstructured' Decision Processes," p. 246-275. *Administrative Science Quarterly*, Vol. 21, 1976.

19. P. Soelberg. "Unprogrammed Decision Making," *Industrial Management Review (renamed Sloan Mgmt Review)*, p. 19-29, Spring, 1967.

20. H. Mintzberg, R. Duru and A. Théorêt. "The Structure of 'Unstructured' Decision Processes," p. 246-275. *Administrative Science Quarterly*, Vol. 21, 1976.

21. R. Bohn. "An Informal Note on Knowledge and How To Use It," *Harvard Business School Working Paper*, 9-686-132, Rev. 4/88, 1986.

22. J. March and H. Simon. Organizations, John Wiley & Sons Pub., 1958.

23. H. Mintzberg, R. Duru and A. Théorêt. "The Structure of 'Unstructured' Decision Processes," p. 246-275. *Administrative Science Quarterly*, Vol. 21, 1976.

24. C. Kiesler and Sproull. "Managerial Responses to Changing Environments: Perspectives on Problem Sensing from Social Cognition," *Administrative Science Quarterly*, p. 548-567, Vol. 27, Dec., 1982.

25. V. Cangelosi and W. Dill. "Organizational Learning: Observations Toward a Theory," *Administrative Science Quarterly*, p. 175-203, Vol

10, Sept., 1965.

26. P. Soelberg. "Unprogrammed Decision Making," *Industrial Management Review (renamed Sloan Mgmt Review)*, p. 19-29, Spring, 1967.

27. P. Soelberg. "Unprogrammed Decision Making," *Industrial Management Review (renamed Sloan Mgmt Review)*, p. 19-29, Spring, 1967.

28. H. Mintzberg, R. Duru and A. Théorêt. "The Structure of 'Unstructured' Decision Processes," p. 246-275. *Administrative Science Quarterly*, Vol. 21, 1976.

29. R. Bohn. "An Informal Note on Knowledge and How To Use It," *Harvard Business School Working Paper*, 9-686-132, Rev. 4/88, 1986.

30. C. Kiesler and Sproull. "Managerial Responses to Changing Environments: Perspectives on Problem Sensing from Social Cognition," *Administrative Science Quarterly*, p. 548-567, Vol. 27, Dec., 1982.

31. C. Kiesler and Sproull. "Managerial Responses to Changing Environments: Perspectives on Problem Sensing from Social Cognition," *Administrative Science Quarterly*, p. 548-567, Vol. 27, Dec., 1982.

32. M. Lyles and I. Mitroff. "Organizational Problem Formulation: An Empirical Study," *Administrative Science Quarterly*, p. 102-119, March, 1980.

33. R. Bohn. "An Informal Note on Knowledge and How To Use It," *Harvard Business School Working Paper*, 9-686-132, Rev. 4/88, 1986.

34. R. Chapman, J. Kennedy, A. Newell, W. Biel. "The Systems Research Laboratory's Air Defense Experiments," *Management Science*, p. 250-269, Vol. 5, 1959.

35. A. Hirschman and C. Lindblom. "Economic Development, Research and Development, Policy Making: Some Converging Views," *Behavioral Science*, p. 211-222, Vol 8, 1962.

36. H. Mintzberg, R. Duru and A. Théorêt. "The Structure of 'Unstructured' Decision Processes," p. 246-275. *Administrative Science Quarterly*, Vol. 21, 1976.

37. C. Kiesler and Sproull. "Managerial Responses to Changing Environments: Perspectives on Problem Sensing from Social Cognition," *Administrative Science Quarterly*, p. 548-567, Vol. 27,

Dec., 1982.

38. R. Bohn. "Management of Learning in Ramp-Ups: Research Summary," Harvard Business School Working Paper, April, 1987.

39. R. Bohn. "Learning by Experimentation in Manufacturing," Harvard Business School Working Paper, May, 1987.

40. R. Bohn. "Noise and Learning in Semiconductor Manufacturing," Harvard Business School Working Paper, May, 1988.

41. M. Lyles and I. Mitroff. "Organizational Problem Formulation: An Empirical Study," *Administrative Science Quarterly*, p. 102-119, March, 1980.

42. R. Bohn. "Noise and Learning in Semiconductor Manufacturing," Harvard Business School Working Paper, May, 1988.

43. M. Tyre. "Managing Innovation in the Manufacturing Environment: Creating Forums for Change on the Factory Floor," M.I.T. Sloan School of Management working paper #3005-89-BPS, December, 1989.

44. M. Tyre. "Managing the Introduction of New Process Technology: An International Comparison," M.I.T. Sloan School of Management working paper, December, 1988.

45. M. Tyre. "Managing Innovation in the Manufacturing Environment: Creating Forums for Change on the Factory Floor," M.I.T. Sloan School of Management working paper #3005-89-BPS, December, 1989.

CHAPTER 5: HYPERCHIP PROBLEM SOLVING ENVIRONMENT

Motorola's Hyperchip TAB ramp-up is not an easy environment on problem solvers. It combines intense challenges with a great deal of uncertainty. Hyperchip managers work under five very different pressures. Their goals are to:

(1) Meet immediate production requirements;

(2) Achieve stability/consistent quality on the production line;

(3) Implement improvements in product design;

(4) Advance TAB technology to be ready for the next generation of TAB products; and,

(5) Establish the infrastructure necessary to have a successful multi-company partnerships in the future.

Tyre points out that the first pressure tends to take precedence over the other four, if for no other reason than that production metrics are usually very clear and feedback is usually relatively rapid[1]. The uncertainties inherent in the ramp-up make it less apparent when problems arise on other dimensions. The bulk of this chapter discusses the sources and types of uncertainty and their impact on problem solving in a high pressure environment*.

5.1) UNCERTAINTY AT HYPERCHIP

The most obvious source of uncertainty in the Hyperchip project is

---

* The illustrations presented in this chapter have been modified to protect proprietary information.

sheer technical complexity. Chapter 1 provided a brief description of TAB. There are one hundred or so steps in the TAB tape making process studied in Chapter 2 and Chapter 3. They each involve a number of process variables — time, temperature, pressure, pH, agitation, humidity, cleanliness, etc.. The bump process and ILB Assembly involve a similar quantity of variables. The range of technical variables is clearly enormous and the number of possible interactions among variables is staggering.

The fact that there are many variables and interactions between variables does not, in itself, foster uncertainty. If every variable and its interactions are well characterized, then the only secret to eliminating uncertainty is a powerful computer. An important source of uncertainty is the level of knowledge about each variable. In an immature technology, decision makers understand little about important interactions among variables and how those interaction impact on the overall system. It is often difficult on detecting a problem to make even Mintzberg's first order cause-and-effect assumptions about obvious variables.

Implicit in the statement that not everybody knows all things about all variables is the understanding that different people know different things. There are two sorts of localized knowledge in organizations. The first is codified knowledge. Formal training is a good example of unshared, codified knowledge. While there is nothing inherently mysterious about the material covered in a formal class, only those who have been exposed to the training will possess its knowledge. The knowledge from formal training tends to be localized because specific organizations need specific sorts of skills. Thus a process engineering group has a relatively high fraction of chemical, materials and mechanical engineers while a circuit design organization has a large number of electrical design engineers. The second sort of localized knowledge is or "uncodifiable" knowledge. This knowledge is the sort that its possessor can not put into words. Most engineering environments involve a good deal of what might be called "art". Lab technicians and operators are often skilled craftsmen — artists in their trade. It is impossible to document all their knowledge in a way to make it accessible to everyone. A crude cut at it is that book learning tends to foster local codified knowledge and experiential learning tends to be the source of local tacit knowledge. At some point of erudition and under high enough time pressures, the effect of local codified knowledge is the same as for uncodifiable knowledge — it is just as mysterious to "neophytes" as an artist's skill. The existence of this localized knowledge, in both of its forms, is a critical source of uncertainty.

A third source of uncertainty in a dynamic environment is that the system constantly changes. Important factors today may be trivial tomorrow and vis a versa. Mintzberg's assertion that opportunities evolve into problems and problems into crises the longer they are left unattended is particularly true in a fast ramp-up. The temporal element is fundamental to how decision makers perceive problems and how they subsequently act on them. Pounds assumption that managers are aware of the status quo and able to judge the difference between it and an ideal situation is not valid in a continually changing environment — there is no status quo. Nor is it a trivial task to judge the difference between reality and ideality in an environment of technical immaturity.

A unique facet of the temporal element is that some problem solving efforts aimed at reducing existing uncertainties inadvertently introduce secondary sources of uncertainty. Two secondary sources at Hyperchip are flow complexity and organizational complexity. Product flow through Motorola's TAB facility has become very complicated as the result of evolutionary changes to the original flow in response specific problems as they were identified.

As for organizational complexity, from the point of view of Hyperchip TAB, there are numerous groups within both DEC and Motorola which have a direct impact on the product. Often, there is no line relationship between groups until very senior levels of management. Each group has developed (or been assigned) its own set of priorities in an effort to target problem solving activity in specific areas. Like the technical variables, the interactions between organizational variables are more important in generating uncertainty than are the variables themselves. Cangelosi and Dill's disjunctive stresses are major elements of the Hyperchip working environment.

One final source of uncertainty (and pressure) unique to the Hyperchip ramp-up is the presence of MIT and the Leaders for

Manufacturing Program. Motorola and DEC have interacted with the LFM Program with a degree of openness unprecedented for both companies on a project of this sensitivity.

The combination of the pressures outlined at the beginning of the chapter and the factors listed above generates three broad types of uncertainty. Technical uncertainty is an incomplete understanding of technical variables, their interactions and their influence. Flow uncertainty is a lack of confidence in estimating when, how many and in what condition parts will be produced. Organizational uncertainty is a lack of clarity about what are global imperatives, how to balance a group's needs against them and how to access the local knowledge of other groups in solving problems. The following section discusses how the various types of uncertainty influence problem solving.

5.2) TECHNICAL UNCERTAINTY

One of the dangers of technical uncertainty is that problems may remain undetected. The level of awareness of critical variables (maturity of the technology) and in whose head that information resides (local knowledge) are critical determinants of whether or not a problem is perceived. These factors also play an important role in influencing whether and how the problem is solved once it is detected. The impact of these sources of uncertainty, as well as the importance of the temporal element, is examined in this section through examples of problem solving activity in the Hyperchip ramp-up. The examples also serve to clarify the nature of technical uncertainty.

5.2.1) TECHNICAL UNCERTAINTY AND PROBLEM FINDING

One of the influences of technical uncertainty in a dynamic environment is that decision makers prioritize Pounds' templates differently than does Pounds, himself. Pounds views Historical Models as being the most useful for problem detection. However, historical models are most appropriate in an environment with well-characterized variables — like the mature industry Pounds studied. In a fast ramp-up, historical templates are useful only in instances of extreme fluctuations, such as yield crashes, because forecasts developed by extrapolating from past performance tend to underestimate the milestones established for the ramp-up.

The Planning Model is more applicable than the Historical. Decision makers notice "problems" if they miss milestones. However, even this template is of little use if engineers and operators close to the project normally know far in advance when milestones will not be met. Lyles and Mitroff suggest that decision makers become aware of problems through informal channels long before being alerted by Pounds' templates. They grow accustomed to working in an environment characterized by "crisis," knowing that they will fall short of planning requirements. Even planning models may be of little practical use in setting day-to-day problem solving grow accustomed to working in an environment characterized by "crisis," knowing that they will fall short of planning requirements. Even planning models may be of little practical use in setting day-to-day problem solving efforts once it is accepted that targets will be routinely missed. One area in which planning models are particularly useful is in targeting mid-term problem solving efforts — missed milestones must eventually be tackled for the product to ship. Slipping from the planning model highlights general areas in need of problem solving. Also, the usefulness of planning models in day-to-day problem solving depends on the specificity of the plan. Extremely fine-grained plans are very useful in targeting problem solving. But well-refined plans create something of a catch-22 — these plans are the ones most likely to be missed. At the operating level, "problems" arising from slipping from planning schedules usually correspond to the time when senior executives far from the day-to-day operation of the project are appraised of the situation[2].

Pounds' other templates, the scientific model and the external model are more useful in Hyperchip TAB. In a dynamic engineering environment, the scientific model routinely plays an extensive role in problem solving. The engineering portions of this thesis are examples of the scientific model at work. However, relatively few day-to-day decisions are premised on extensive proactive experimentation. Usually, the role of the scientific model is limited to supporting periodic large changes in basic assumptions or processes. External models are by far the most important templates for detecting problems. External models range from the direct influence of the customer through quality specifications to the somewhat trite example above of planning problems not being "problems" until senior executives learned of them. The role of external models is discussed more completely in §5.4: Organizational Uncertainty. The example of specifications has an important influence of technical uncertainty, so it is discussed in this section, also.

Motorola lives by two credos: "Six Sigma*" and "Total Customer Satisfaction". While adherence to these two mottos has earned Motorola the first Malcolm Baldridge Award, it has also complicated the volume delivery of Hyperchip product to Digital. Much of the technology simply does not yet exist to reach six-sigma level consistency or to meet either the extensive customer specifications. Pounds' templates for detecting problems would indicate that failure to meet either goal would signify a technical problem. However, failure at such a global level does not provide much direction for problem solving efforts. One way to target local problem solving efforts is to use individual quality specifications and screen for areas which fall short of them.

One of the difficulties associated with technical complexity and immaturity is that it is difficult to determine precisely what are proper metrics for measuring quality and at what level those metrics should be set.

---

* Six standard deviations of product properties are within specification limits. This requirement translates into a defect rate of about three parts per million.

Where reliability is as critical, as its is for HiEnd, specification writers will tend to err on the side of caution. Occasionally, overly-stringent specifications can be the source of increased uncertainty rather than the solution to it. This situation occurs when a process is modified so that specific product attributes fall within quality and secondary effects of the process modifications generate either misinformation or directly introduce more uncertainty in another area. In an environment of technical complexity, the impacts of process changes are almost never completely understood before the change is implemented. Therefore the risk of secondary uncertainties arising from process changes is particularly high. Local solutions to problems can ripple through the operations and show up as problems in other areas. In many situations, process changes are unpopular because any change risks "losing the recipe."

An additional point here is that in pursuit of long term reliability, specification writers tend to eliminate any potentially negative factor at the expense of short term throughput. In the early stages of a ramp-up, throughput and experiential learning are arguably more important than tight quality as called out in specifications[3]. This issue is revisited later in this chapter.

5.2.2) TECHNICAL IMMATURITY AND PROBLEM SOLVING

One way to cope with technical immaturity is to set specification limits so tight as to eliminate any possibility of a reliability failure and then to use these specifications in detecting problems. This philosophy clearly has some drawbacks, some of which are noted above and one which is discussed more in the following section. A less disruptive way to identify variables is to perform "natural" experiments - either to passively monitor the existing process or to analyze an extensive library of existing data[4]. The lead pull work presented in Chapter 3 is an example of a large natural experiment:

> A priori, it was only known that ILB yield seemed to vary with leveling agent. However, the relationship was not completely characterized. It was not justifiable to initiate a large new experiment simply to detect a problem. But, using the large pool of lead pull data from ongoing engineering experiments and incorporating new data as it was generated, it was possible to show the yield trends discussed in Chapter 3, thereby targeting subsequent problem solving activities.

Even this "natural experiment" was very long and engineering-intensive. The result of weeks of number-crunching was that a problem was defined — not that a solution was discovered. All this effort is an indication that Mintzberg's memory search is not necessarily the easiest (or the first) search technique typically used in problem solving in a technically immature environment.

Another aspect of technical immaturity is that once technical problems evidence themselves, engineers generally have had little prior experience with similar problems. There are very few "off-the-shelf" solutions. Each situation requires one of Mintzberg's custom or semi-custom solutions. Simple answers are so infrequent in a technically immature ramp-up that often the operational definition of a "problem" is a situation which requires a previously untried solution. Fixes are very engineering intensive. Technical uncertainty also tends to make any investigation more complicated than it initially appears. Cursory investigations evolve into comprehensive surveys as engineers think of additional variables which could play an important role in their studies. Engineers get caught up in the cycle of feedback loops shown in all the models in Chapter 4. An example from Chapter 2 and Chapter 3 is a thermode profile study that was initiated as part of the ILB lead pull yield investigation:

> ILB yield surveys were seen as a way of confirming initial suspicions of lead thickness trends correlated to leveling agent. However, soon it became apparent that the condition of the thermode played a critical role in ILB yield. To separate the thermode variable from TAB tape processing effects, an investigation of thermode profiles was initiated. This study proved to be a large engineering task independent of the lead pull study, disrupting the normal bonder setup procedure and adding tasks for TAB Development technicians.

In a complex problem it is difficult to remain focused on one theme. Engineers may be sidetracked into any number of secondary issues, each of which could easily consume all of their time. The practice of profiling thermodes, for instance, could become an integral part of process control on thermode use. The effort to introduce thermode process control could (and did) demand as much engineering effort as the original lead pull investigation.

5.2.3) LOCAL KNOWLEDGE AND PROBLEM SOLVING

One aspect of technical immaturity on a complex product is that most knowledge is tied up in the heads of individuals rather than institutionalized in specifications. Someone has to write specifications. An important variable is who writes the specifications. For instance, in an advanced electronics environment, it might seem like the best people for writing specifications would be electrical engineers trained in circuit design. However, the challenges in the Motorola TAB ramp-up are chemical/physical/mechanical in nature. Therefore, the "best" people to write specifications are physicists and chemists experienced in manufacturing. When formal specifications fails to address the most important manufacturing problems, engineers close to the production operations will rely heavily on personal engineering judgement in detecting problems. They will begin to depend on "temporary" local specifications for daily quality gates. Cangelosi and Dill suggest that these temporary decision rules which are frequently used tend to become permanent benchmarks which precedence over formal quality specifications.

In her discussion of how local information complicates problem solving under conditions of technical uncertainty, Tyre notes that workers make assumptions about reality and then unilaterally act on these assumptions[5]. Similarly, Mintzberg notes (in Chapter 4) that decision makers make first order approximations between obvious variables based on their level of understanding. Without training in specific codified knowledge, operators occasionally make incorrect assumptions. The result can be that operators may inadvertently introduce larger problems than the ones they attempt to solve. Similarly, operators tend to become craftsmen in their particular tasks. This skill is largely uncodifiable knowledge of the individual operators. Attempting to put that knowledge to paper and transfer it as "book learning" may also inadvertently introduce secondary uncertainty. Bohn suggests that this situation is typical of complex ramp-ups[6,7].

5.2.4) KIESLER & SPROULL AND ENGINEERING JUDGEMENT

Given the tremendous number of potential variables to study, Hyperchip engineers often use engineering judgement in selecting appropriate variables and in making final decisions. "Engineering judgement" may be defined using Kiesler and Sproull's terminology presented in Chapter 4. However, rather than being "interpretive pitfalls", the shortcuts that decision makers use to asses the validity of information can be powerful tools in a complex ramp-up. The reason is that it is impossible to objectively weight information when as much information confronts engineers as they do in the Hyperchip environment. Decision makers simply do not have the time to completely characterize every problem confronting them.

The danger comes when decision makers use the shortcuts without realizing they are using them. Then they are uniquely vulnerable to misinterpreting data they incorporate. Conversely, they are also vulnerable to interpretive pitfalls when they fail to use shortcuts due to Kiesler & Sproull's interpretive biases. An example of an opportunity where shortcuts could have been used but were not is the entire engineering investigation outlined in Chapter 2 and Chapter 3:

> The motivation for both investigations was to quantitatively confirm the engineering judgement of one senior TAB development engineer. He needed "hard" data to support his assertion that ILB yield was related to leveling agent. Other engineers resisted his assertions without "proof". Months of number crunching subsequently bore out his initial hunch. On the basis of that investigation, several suggestions were made for potential modifications to the TAB tape manufacturing process.

Decision makers could have accepted the engineer's opinion and modified the TAB process on the strength of his credibility. They missed the opportunity to do so. Instead, they devoted considerable time and engineering resources to "proving" his assumption rather than dedicating them to other problems. In retrospect, they fell prey to Kiesler & Sproull's pitfalls twice: once by not using the shortcut and once by not accepting information as valid because it was not "schema-relevant". They resisted his suggestion because it did not fit into pre-established notions of TAB tape dimensional tolerances.

5.2.5) TEMPORAL UNCERTAINTY AND TECHNICAL PROBLEMS

One of the most confusing aspects of technical uncertainty is the influence of the temporal factor. Perhaps the most frustrating phenomenon is the problem that solves itself:

> During a tape steering committee meeting, one engineer noted the tape yield was unusually high. Another engineer asked what the source of the improvement had been. No one knew — the technology is immature; sometimes things change without explanation. "And that's what worries me," replied the cynic, "If you don't know what got you there, it could go at any time." Later, the tape yield fell and, similarly, no one had an explanation.

Why did the problem go away? Will it reemerge as mysteriously as it disappeared? Has it been resolved at the expense of a another variable and the trouble not yet detected?

Another aspect of the temporal element is that priorities constantly shift as minor inconveniences become major dilemmas or as old problems are solved and new problems take their place on the top of the pareto chart. One of Kiesler and Sproull's "interpretive pitfalls" listed in Chapter 4 is that decision makers tend to pay attention to only the largest signal. Problem detection by pareto chart is an example of this sort of behavior — problem solvers concentrate only on the largest columns on the chart and worry about the other things later. They miss opportunities to simultaneously tackle similar problems which may be separate on the pareto chart. Or they recognize where future problems will occur, but decline to address them before they have reached the top of the chart. One example of this sort of behavior is chasing bottlenecks through an operation, solving them one-at-a-time as they shift from one operation to the next. Another common practice is leaving problems only partially resolved as soon as they fall from the top spots on the chart. Early problem solving activities tend to mitigate problems just enough to expose other problems.

Another feature of the temporal element is that managers sometimes wait for solutions to manifest themselves in the natural course of events (they sometimes do), rather than investing precious resources in attacking them immediately. This tendency is most common when new equipment or new process changes are scheduled. It is potentially the most visible example of Mintzberg's Passive Search in s ramp-up:

> One common belief about the TAB Assembly process was that a number of challenges would be eliminated as a result of a reorganization in which the Production operation would be split from the Development operation. However, some issues which impacted directly on ILB yield were traceable either to processes not scheduled to be moved or to processes to be moved unchanged to the new facility. The new clean room would not

directly address these causes.

Problem solving activity does not decrease while a passive search is ongoing. But, in this situation, certain problem solving efforts were put on hold pending the introduction of new facilities. There seemed to be no point in solving problems twice — once "now" and once again when the new facilities were established. The danger in this approach is in the possibility of entrenching poor practices in new infrastructures. Old problems may overshadow improvements brought by innovation. Moreover, new techniques may be more sensitive to fundamental problems than old methods, which have already been made somewhat robust to underlying problems. Decision makers must be rigorous in their estimation of problems which are associated only with old processes and equipment before assuming they will be remedied by innovations. In an environment of high technical uncertainty, this estimation is very difficult to make with any degree of confidence.

5.3) PRODUCTION FLOW UNCERTAINTY

Production flow uncertainty stems from three sources: (1) Inherent yield and cycle time uncertainties at individual operations; (2) Incoming product quality and volume variability; and, (3) Confusing materials flow patterns. Bohn stresses that in any ramp-up, there will be a high degree of uncertainty associated with incoming materials quality and with individual operations[8,9,10]. These two types of uncertainty may be roughly grouped as "conversion" uncertainty — they are inherent in the starting materials or the processes used to convert starting materials into final products. Hyperchip is a typical of other ramp-ups with respect to its conversion uncertainty — except for wafer bump (developed entirely on-site at Motorola's TAB facility and, thus very well characterized). The third source of uncertainty might be termed non-operation uncertainty. As suggested in §5.1, the combined influence of evolutionary changes in process flow to alleviate other forms of uncertainty is a primary source of non-operation uncertainty for Hyperchip. Because conversion uncertainty is largely a function of technological maturity requiring hard-core engineering fixes, it is important minimize non-operation sources of uncertainty using managerial fixes so that engineers may focus on advancing the technology.

As Motorola's TAB production flow has evolved into a complicated pattern, they have come to rely increasingly on expediters for their materials handling. The variation in the rate at which the tasks are carried out at each station adds to the complexity. Physical product and paperwork occasionally arrive at different times. One of the most difficult jobs is that of the engineer whose duty it is to monitor each chip and tape site as expediters hand-carry Motorola's TAB facility. Her job is to make the unshared knowledge of the expediters available to DEC an Motorola engineers as well as to verify its validity and to forecast short term throughput using that information.

To cut cycles time and minimize uncertainty, Motorola recognizes the need to simplify their production flow. However, the immediate demands of the ramp-up provide engineers little time to step back and look at the "big pictures". Because this system evolved incrementally, it is sometimes easy for the direct players to lose sight of the forest for the trees:

> A day shift production engineer was asked to generate a plan for increasing throughput by about 100%. One suggestion was to reduce in the scope of a quality assurance gate whose length was not a bottleneck (though the amount of time parts sat in the QA queue prior to the inspection sometimes was). The other major suggestion was to add expediters — more people to hand-carry lots to increase throughput rather than a simpler production flow to increase throughput.

To minimize flow uncertainty from upstream operations, Motorola's standard practice is to implement Process Change Control. They freeze upstream processes and insist on buy-in from downstream engineers for changes to the process. This practice prevents upstream operations from making ad hoc procedural changes which impact in unforeseeable ways on subsequent operations. While any effort to decrease uncertainty would seem good, in the early phases of a ramp-up, Process Change Control may not be worth the bureaucracy it adds to process innovation. As one engineer explained it when Motorola initially considered placing an upstream operation under Process Change Control:

> "This stuff works for operations that already have a mature process. This one isn't, yet. We don't know what changes are

> heavy hitters. But we do know that we are going to need the latitude to make changes between now and volume production. Process Change Control will take away that latitude."

At about the same time Motorola considered implementing Process Change Control on certain upstream operations, DEC and Motorola eliminated much of the bureaucracy stifling downstream learning:

> In the early phase of the production ramp-up, tight quality specifications that were written to ensure DEC's OLB assembly facility received nothing but quality parts slowed total delivery of parts to a trickle. Instead of permitting concurrent engineering, the volume of parts forced Motorola and DEC to attack problems strictly in series. Motorola could not solve its assembly problems until enough components passed upstream gates to expose problems in the TAB assembly process. DEC could not troubleshoot its own problems until enough of Motorola's assembled product passed quality gates to highlight their own problems. To address this dilemma, the companies agreed to temporarily shift in-process gates to the end of the entire TAB Assembly process as one Final Outgoing Inspection. Motorola would "reject" parts according to the stringent specifications. Then a DEC inspector would then use his engineering judgement to determine which rejections were legitimately "bad" parts and which were artifacts of the specifications that could be shipped to DEC's OLB facility.

This situation highlights an important aspect of the Hyperchip ramp-up: immediate Total Customer Satisfaction and long term Six-Sigma Quality are not necessarily compatible goals. Once throughput increased, both companies were able to substantially improve their operations and increase overall product quality and process robustness. This observation runs counter to current fashion that "quality is free." In the early stages of a ramp-up, the price for unreasonably high quality is throughput.

5.4) ORGANIZATIONAL UNCERTAINTY

Most hurdles to solving the technical and flow problems described above are communications issues between groups. Cangelosi and Dill feel that disjunctive stress between groups plays the largest role in defining problems.[11]. Similarly, Pounds feels that External Models are second only to Panning Models in defining problems — and even then, slipping from planning models represents a problem only to the extent external agents such as senior management react negatively to it[12]. Mintzberg notes that the chief source of information is through verbal communication with external parties rather than reference to existing storehouses[13].

The parallel nature of the organizations in Hyperchip TAB, both within Motorola and across company boundaries, means that decisions are necessarily premised on a great deal of networking and buy-in. Both DEC and Motorola are accustomed to operating in this mode. DEC's culture is based almost exclusively on networking and buy-in. Motorola has gone to great lengths over the past five years to infuse participative management throughout the company. Conventional wisdom on consensus building is that it takes a long time to get all interested parties to buy into a decisions, but that once everyone is on-board, the execution of the decision is very rapid and successful because everyone shares the same vision.

5.4.1) MEETINGS

To minimize the opportunities for mis-communication between groups, Motorola and DEC hold frequent meetings between players at all levels of the project. Executives and senior engineers associated with the project meet periodically to make sure everyone is "rowing in the same direction." Motorola and DEC engineers frequently meet amongst themselves and with engineers from critical vendors to focus their problem solving efforts.

All these meetings serve both the constructive purposes of disseminating unshared information and the not-so-constructive purposes of acting as a medium for Soelberg's confirmation process. In an immature technology little is known about the extended consequences of important decisions. Unfortunately, in a ramp-up with an emphasis on forward progress it is very difficult to "take back" any decision. The people making day-to-day decisions in Hyperchip are in a position very similar to the to one in which Soelberg's job-seekers found themselves. They constantly face the task of making irreversible, open-ended choices. The stress associated with making these sorts of decisions makes this environment susceptible to Soelberg's confirmation process. However, in a ramp-up with a premium on speed, there is not time to build an active roster of more than one solution to any particular problem. Decision makers have to make a choice without the benefit of a confirmation candidate to act as a straw man. They therefore sometimes use meetings as a proxy for a confirmation candidate. In all likelihood, they are not aware that they are engaging in a confirmation process. They may feel they are working on consensus-building. However, sometimes buy-in is not true consensus building but endorsement of a risky position.

5.4.2) CONFLICTING PRIORITIES

Perhaps the most obvious interface between organizations is the one between Motorola and DEC. Both companies have tried to make this junction as seamless as possible. Several DEC engineers are periodic residents at Motorola's TAB facility. They are integrated into the day-to-day operation at Motorola. Most Motorolans see the DEC engineers as critical members of the project team. DEC engineers try to act as Hyperchip team resources rather than as Digital "cops" monitoring a vendor's operations for conformance to quality standards.

Some of the tensions affecting the daily operation of Hyperchip TAB arise between the groups with different responsibilities for the TAB project. One tension between groups which play an important role in Hyperchip

TAB but which also have responsibilities outside TAB. Internal to the Hyperchip TAB effort, divergent priorities between groups sometimes result in the adoption of non-optimal procedures. The priorities of one group may dominate to the detriment of other aspects of the effort:

> Motorola experienced periodic difficulty probing leads on certain tape during their test-on-tape operation. The tape was modified to be testable by adding steps to the front end of the tape fabrication process. The test problem was at the back end of the entire ILB assembly operation. Thus, the solution represented a major change at the earliest possible location in the combined tape fabrication/ILB assembly process. An early change has an opportunity to ripple through all subsequent processes. In a technologically immature arena, the ultimate influence of such a change is to add uncertainty.

5.4.3) NOT INVENTED HERE

In any large project a significant barrier to collaboration between the groups is not simple mis-communication or conflicting priorities. Even with the high degree of cooperation between companies and the team spirit on the Hyperchip TAB Project, the "Not Invented Here" syndrome (NIH) is an obstacle to the definition and resolution of problems. There are three sources of NIH at Hyperchip: (1) Pride in the accomplishments of the parent organization; (2) Distrust of solutions developed external to the organization; and, (3) Fear on the part of upstream vendors either of vertical integration by Motorola or DEC into their product line or concern that they might inadvertently leak secrets to potential competitors.

The Hyperchip TAB Development team at Motorola is justifiably proud of its position at the leading edge of TAB technology. They are, conversely, suspicious of "breakthroughs" in the technical community and in other areas of both Motorola and DEC. Recent advances outside their organization often represent hurdles they have already had to overcome. Being pioneers in the field, their organization is often the only place where much of the specific knowledge on high-density TAB exists. Like Kiesler and Sproull's "Interpretive Pitfalls," not all NIH is necessarily undesirable. NIH may serve as a useful screening technique for time-constrained decision makers in much the same way that Kiesler and Sproull's shortcuts may be used.

NIH also shows up as resistance to change on the production line. It has taken a significant effort to get a process to consistently manufacture Hyperchip TAB packages. There is little known about the extended repercussions of changes on the line, so there is a strong undercurrent of "don't tamper with success." When engineers "impose" changes on the line, they add work for the operators (at least temporarily) and risk "losing the recipe". Production workers tend to be suspicious of promises of better results some time in the distant future that immediately disrupt that stability. The "prove-it-to-me" stance which prompted the time-consuming lead pull investigation detailed earlier is an example of this aversion to change. Immediately accepting the opinion of the senior engineer about leveling agents might have meant fundamental changes in the tape fabrication operation which risked losing the recipe. This situation is an example of where NIH might have a positive aspect: the time taken to quantitatively characterize the extent of the problem gave the tape operation an opportunity to phase-in a change without disrupting their line.

One of the most significant lost opportunities is failure on the part of critical vendors to recognize that they have the tremendous combined analytical resources of Motorola and DEC at their disposal in solving their internal technical problems. Very rarely does a situation arise where the success of a project is as important to the mid-term strategy to companies as Hyperchip TAB and HiEnd are to Motorola and DEC. Vendors in the critical supply chain are central to that success. Vendors have a "once-in-a-life-time" opportunity to utilize DEC and Motorola resources in their own problem-solving efforts. While critical vendors are aware of the immediate support of Motorola's TAB Development group and DEC's parallel organization, they are perhaps not as aware that they can tap into the extended facilities of both companies. An example of the priority Motorola places on the Hyperchip TAB project is the immediate attention that SPS's ESCA and Auger microscopy laboratories were willing to devote to Hyperchip TAB during several investigation. They routinely placed Hyperchip at the front of long lines of backlogged work. Most likely, the vendors have largely passed on this opportunity out of concern for revealing proprietary secrets to the competition.

CHAPTER 5: HYPERCHIP PROBLEM SOLVING ENVIRONMENT

5.5) SUMMARY

This chapter outlines uncertainties which exist at Hyperchip and in ramp-ups in general. It illustrates the influence of those uncertainties on decision making. Since many sources of uncertainty are inherent in a technically complex ramp-up, it is important to minimize uncertainties which are introduced by localized, evolutionary management practices. It is also critical to minimize the barriers to communication between groups to maximize the flow of specific information to problem areas. The next chapter deals with some potential ways to meet these needs. Much of the solution must come from senior executives — middle level decision makers work very hard at Hyperchip to optimize their pieces of the project. It is up to senior management to provide the linkages between organizations and the direction necessary to make the combined resources of both companies available to solve problems in a coordinated fashion.

CHAPTER 5: HYPERCHIP PROBLEM SOLVING ENVIRONMENT

1. M. Tyre. "Managing Innovation in the Manufacturing Environment: Creating Forums for Change on the Factory Floor," M.I.T. Sloan School of Management working paper #3005-89-BPS, December, 1989.
2. H. Mintzberg, R. Duru and A. Théorêt. "The Structure of 'Unstructured' Decision Processes," p. 246-275. *Administrative Science Quarterly*, Vol. 21, 1976.
3. R. Bohn. "Noise and Learning in Semiconductor Manufacturing," Harvard Business School Working Paper, May, 1988.
4. R. Bohn. "Management of Learning in Ramp-Ups: Research Summary," Harvard Business School Working Paper, April, 1987.
5. M. Tyre. "Managing Innovation in the Manufacturing Environment: Creating Forums for Change on the Factory Floor," M.I.T. Sloan School of Management working paper #3005-89-BPS, December, 1989.
6. R. Bohn. "Learning by Experimentation in Manufacturing," Harvard Business School Working Paper, May, 1987.
7. R. Bohn. "Noise and Learning in Semiconductor Manufacturing," Harvard Business School Working Paper, May, 1988.
8. R. Bohn. "Management of Learning in Ramp-Ups: Research Summary," Harvard Business School Working Paper, April, 1987.
9. R. Bohn. "Learning by Experimentation in Manufacturing," Harvard Business School Working Paper, May, 1987.
10. R. Bohn. "Noise and Learning in Semiconductor Manufacturing," Harvard Business School Working Paper, May, 1988.
11. V. Cangelosi and W. Dill. "Organizational Learning: Observations Toward a Theory," *Administrative Science Quarterly*, p. 175-203, Vol 10, Sept., 1965.
12. W. Pounds. "The Process of Problem Finding," *Industrial Management Review (renamed Sloan Management Review)*, p. 1-19, Fall, 1969.
13. H. Mintzberg, R. Duru and A. Théorêt. "The Structure of 'Unstructured' Decision Processes," p. 246-275. *Administrative Science Quarterly*, Vol. 21, 1976.

CHAPTER 6: CONCLUSIONS

There are two general sets of conclusions in this chapter. The first set is largely a summary of some of the points made in Chapter 5 about the insights the cognitive models offer in a ramp-up environment. The second section presents recommendations on improving the decision environment at Motorola.

6.1) APPLICABILITY OF MODELS TO RAMP-UP ENVIRONMENTS

6.1.1) PROBLEM RECOGNITION

In a ramp-up in an immature technology, problems are rarely recognized on the basis of Pounds' historical models (detecting problems on the basis of comparison with past performance). More often, they are noticed because either performance fails to meet ramp-up goals (Soelberg's "discovery of a barrier to progress toward an objective") or because performance fails to meet specification limits. Soelberg's contention that decision makers must have some sense of ownership of a problem is particularly true in a ramp-up where problem solvers have far too many items on their agenda to pay enough attention to any one of them. Only in rare cases are problems proactively detected in a ramp-up — there are enough problems which "volunteer" themselves to keep engineers occupied.

Problem solving activity in a ramp-up is motivated far more often by the threat of failure than by Chapman's "discomfort" stress. Failure is usually operationally defined using Pounds' templates either by slipping from ramp-up milestones (Planning Model) or by failing to conform to specifications (External Model). Tensions between various organization also play a large role in the form of Cangelosi & Dill's "disjunctive" stress in motivating problem-solving efforts. Most problems in a ramp-up are what Mintzberg would term crises — Mintzberg's "opportunities" almost never generate problem solving activities in a ramp-up where engineering resources are dedicated to firefighting. Bounding the decision space is often very difficult for complex problems as additional variables are discovered (or imagined) which might impact on the solution. Chapter 5 discussed the thermode profiling activities as an extension of the original definition of the lead thickness variation problem.

Of critical importance in an environment where there is a large amount of local, unshared information is the issue of who defines the problem. With incomplete information, whoever defines a problem will likely make mistakes estimating Mintzberg's "first-order interactions between obvious variables" unless they are privy to the specific knowledge which directly addresses the problem under consideration.

6.1.2) SEARCH

Decision makers in a ramp-up do not necessarily follow in order the steps Mintzberg outlines for a Search. Mintzberg believes that a search begins with a quick review of organizational memory and progressively becomes more active as it fails to generate minimally sufficient solutions. His hierarchy is Memory Search, Passive Search, Trap Search, and Active Search.

The ease of a memory scan depends greatly on the form in which the data is stored. Compiling and interpreting the data for Chapter 3 was very engineering intensive — much more so than entire custom solutions for other problems. Organizational memory is particularly difficult to access when a large amount of the knowledge is tacit, or "uncodifiable," as it is in Hyperchip. In a typical dynamic situation, it is almost assumed that organizational memory is not enough information to address the problem, so problem solvers skip directly to higher order search techniques.

In Mintzberg's model, Passive Search occurs sequentially after Memory Scan fails. In general, there is not time in a ramp-up for this sort of search routine. When it does occur, it will typically manifest itself as waiting for new equipment or new process changes to solve problems almost incidentally.

One aspect of Mintzberg's hierarchy is almost never employed at Hyperchip. Between the delay waiting for external solutions and the distrust of solutions once they trickle in, the "trap search" is almost never employed. Any situation which might merit a "trap search" most likely would not be considered a problem, but an inconvenience.

CHAPTER 6: CONCLUSIONS

The most common form of Search in a ramp-up (especially in the early phases) is an Active Search — i.e. experimentation. Since the technology is immature, there will be little organizational memory to scan. The time constraints of most ramp-ups preclude most passive search routines. Time constraints and distrust of external results combine to discourage trap searches. Also, many problems is a ramp-up ultimately require custom solutions — i.e. previously untried solutions highly modified solutions. Knowledge gained through active search can directly aid in the formation of a custom solution.

Mintzberg points out that custom solutions are so resource intensive that normally only one custom solution is formulated before the search is stopped. In a ramp-up on a technologically immature product, most problems require custom solutions. Furthermore, the time constraints are so tight that even in situations where there are "ready-made" solutions, the search is generally stopped after the first satisfactory solution is found. The active roster almost <u>never</u> gets to be larger than one.

March and Simon's idea of seeking a satisfactory solution rather than the optimum solution is particularly applicable in a ramp-up. Occasionally, however, it is very difficult to find even minimally acceptable solutions in the time allowed by the ramp-up. In this situation, one of the most maligned but useful options available to the decision maker is to redefine "minimally acceptable". Often in a technically immature technology, the original definition of what is satisfactory is based on incomplete understanding all variables. As implications of initial standards become clear, it makes sense to re-evaluate those criteria. Motorola and DEC reexamined how and when they evaluated "quality" when they temporarily shifted certain quality gates to the end of the TAB Assembly operation to increase throughput to DEC OLB.

6.1.3) SELECTION

March and Simon presented the idea of satisficing behavior as a more realistic mechanism for problem solving than optimizing. Their point was that it is prohibitively difficult to optimize in any realistic problem situation. Their idea is particularly applicable in a complex ramp-up where time pressure and uncertainty make it very difficult to consider any more than one satisfactory alternative, let alone enough alternatives to find the "optimal" solution.

Soelberg proposes the confirmation process. Decision makers forced to make open-ended, irreversible choices experience anxiety when they do so — especially if the consequences of the decision may be very large. Soelberg suggests that problem solvers never explicitly select a "best" solution without having a straw-man solution against which to contrast it. In a fast ramp-up, there is not enough time to develop such a confirmation candidate. Decision makers use other mechanisms to confirm their favorite choices. One common mechanism is endorsement of other decision makers and engineers during meetings.

6.1.4) ENGINEERING JUDGEMENT

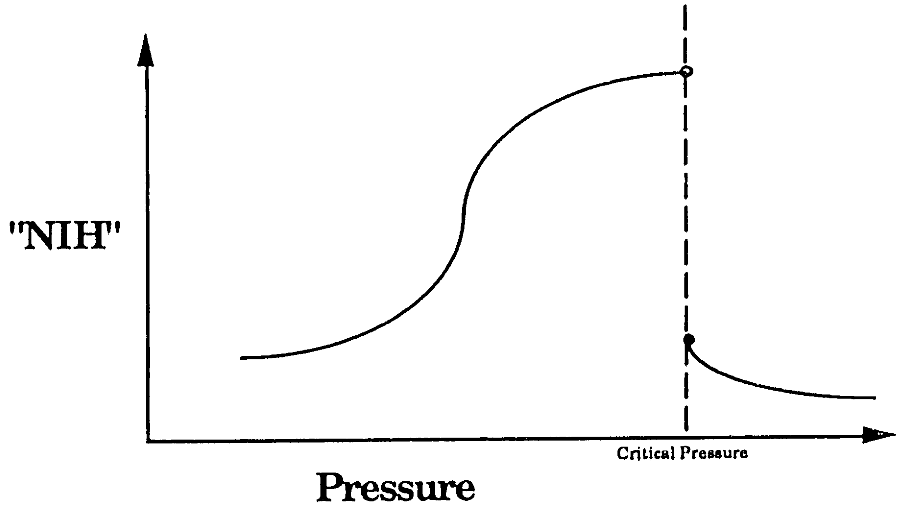
Kiesler and Sproull's "interpretive pitfalls" might also be called "shortcuts for evaluating the relevance of information." They are only pitfalls if decision makers misuse them or fail to use them given the opportunity. The practical distinction between good engineering judgement and mis-interpreting data is largely a matter of experience. Since engineering judgement is a very large aspect of any ramp-up in an immature technology, a corollary is that there should be experienced engineers involved in the daily decisions in a ramp-up.

6.1.5) LEARNING

Solutions implemented without sufficient understanding have a good chance of introducing added uncertainty into the environment. The need to learn is a reflection of the temporal element present in a dynamic environment. Bohn recognizes that the most important managerial priority in a ramp-up is the management of learning. It is also the most difficult priority to keep sight of. The immediate needs of production almost always overshadow the need to learn. In Hyperchip, the most obvious example of this skew is in the TAB Development groups, whose charter is to introduce next-generation TAB, but whose day-to-day efforts are devoted almost exclusively to the immediate problems of Hyperchip production. In as much as volume is an important aspect of uncovering process problems, throughput is an important aspect of learning. As shown in Chapter 5, short term quality requirements which minimize throughput stifle learning in the early phases of a ramp-up.

Perhaps the largest barrier to organizational learning and the dissemination of specific/local knowledge is the "Not Invented Here" syndrome. NIH is typical of an environment with multiple organizations, high pressure, technical immaturity and tacit knowledge. NIH increases with pressure in a technically immature environment. The reason is that the consequences of failure increase with increasing pressure. Therefore the premium on being certain the decision is a correct one increases. The best way to be confident in a solution is to have developed it internally. If pressure continues to increase, at some critical threshold, the organization will throw in the towel and seek outside help. Figure 6.1 illustrates schematically how NIH varies with pressure:

Figure 6.1: The NIH Response to Pressure



6.2) RECOMMENDATIONS FOR HYPERCHIP

Chapter 5 discussed how the pressures and uncertainties of the Hyperchip ramp-up impact on the problem solving process there. There are three possible steps which would make problem solving less burdensome in Hyperchip: (1) Add more problem-solving resources; (2) Eliminate the uncertainties; and, (3) Insulate the problems solvers from the uncertainties. The first parts of this section discusses several specific steps that may be taken at Hyperchip to address specific uncertainties. The point of these parts is to illustrate that once specific problems are defined, outlining specific solutions is a relatively simple task. The final, and perhaps most important section of the chapter discusses the dual roles of senior management to both insulate the operating levels from excess pressure and to bring diverse resources to bear on Hyperchip problem solving efforts.

6.2.1) ADD ENGINEERS

The technical complexity of the Hyperchip Project is a fixed variable. Management fixes may affect it only indirectly. The only way it may be reduced directly is through engineering effort. Therefore, one change that immediately suggests itself is to bring more engineering support to the project.

To avoid inadvertently increasing complexity by randomly throwing more people into the situation, it is important to recognize what types of engineering support are important for Hyperchip TAB. Few of the challenges are traditional electrical engineering problems any longer. The majority of the issues which face Hyperchip are electrochemistry, metallurgy, polymer or heat transfer problems. The types of engineers needed are people with backgrounds in those areas. Also, to minimize the disruption that occurs as engineers get "settled-in," the engineers brought into the project should ideally be experienced in electronic packaging and have several years experience with the company.

6.2.2) ELIMINATE EXTRINSIC, SECONDARY UNCERTAINTIES

Because of the technological complexity and immaturity of Hyperchip TAB and HiEnd, there is a high degree of intrinsic uncertainty that is hard to eliminate. It is all the more important, therefore, to minimize uncertainty arising from external pressures or from evolutionary changes. Some steps have already been taken in this direction at Motorola's TAB facility. One example is the decision to shift quality gates in order to increase throughput to DEC's OLB Facility. In this situation, increased consistency in the volume of product shipped to the customer was more important in the developmental stages of the ramp-up than was absolute product quality.

There are other, relatively straightforward actions that could be taken. One is to simplify the production flow through Motorola's TAB facility. One simple change that might be made with minimal disruption is to add more "windows" into clean room areas through which parts may be moved. Currently, expediters dress and undress multiple times as they carry parts form one clean room to another.

Another way to limit uncertainty is to spread local knowledge. One area where specific information needs to be shared at Hyperchip is from line operations to engineers and from engineers to line operators. One way to accomplish this sharing is to rotate engineers through operations as production engineers and line operators through engineering in meetings and as technicians. As it stands today, engineers with extensive formal training have relatively little direct exposure to production activities. Their schedules do not permit them to spend time on the line. Line operators should also be exposed to some of the reasons behind what they are doing. For instance, an operator should always be involved in the daily TAB Task Force meeting. Operators are currently about the only people who are not directly involved in the meeting. Their presence would both increase the accuracy of the information the TAB coordinator gets and provide line operators with a better understanding of global imperatives.

6.2.3) INSULATE PROBLEM SOLVERS

Tyre suggests to partially insulate the introduction of new solutions from the day-to-day demands of production by using "forums for change." In one example she cites, these forums were physically separated from the daily production pressure. The two critical factors in a forum for change are this separation from production pressures and close proximity to the production environment. Separation from production pressures allows innovations to be measured against independent metrics from production practices. It also protects against the possibility that development tools will be pressed into production service.

Motorola development engineering first had close proximity to the production environment and then had separation from production pressures, but not both. Initially, the engineering lab and the Hyperchip TAB pilot line were housed in the same clean room. The immediate needs of production sometimes precluded development work. In addition, the pilot-line room was very small for production operators, development engineers and development technicians to all crowd in together. Then Hyperchip production moved into a separate room. Suddenly there was room for experiments. Development staff could perform experiments uninterrupted by production imperatives. However, the physical isolation of the engineering lab from the production environment now made it more difficult to transfer improvements to the production process.

One of Tyre's points is that problem solving efforts get overshadowed by the immediate needs of production when their success is measured using production benchmarks[1]. In Hyperchip TAB, process improvements should constantly be made. A chief difficulty in implementing these improvements is that they risk temporarily hurting production yields. The problem is that the performance of line operations is judged on the day-to-day delivery of product (other benchmarks, such as quality, are binary: it's there or it isn't). This performance metric is incompatible with continual innovation for long term improvement. Motorola is faced with the dilemma of establishing performance metrics which both measure throughput to the customer and progress toward ramp-up objectives. One way to achieve this goal is to spread the measurement over a longer period. That way, temporary perturbations in throughput would be smoothed as innovations resulted in throughput improvements. This issue is discussed further in the next section.

6.3) CONCLUSIONS

Many of the suggestions outlined above must necessarily be driven from the top-down. Some may involve changes in incentive structures — to focus more on process control instead of volume, for instance. Senior managers might also influence line personnel in overcoming their fear that top-down changes risk "losing the recipe" by establishing more of a routine presence in the production environment — i.e. by being more active in "Management By Walking Around." Other activities involve forging linkages across diverse organizations — a task for which the senior managers involved in the periodic review meetings are eminently suited.

CHAPTER 6: CONCLUSIONS

The HiEnd development effort and Hyperchip TAB are an immense experiment in corporate teamwork. To a certain extent, the tone of this thesis reflects the day-to-day pressures and frustrations of a complex ramp-up. It does not properly convey the enormous success of the project. It is a testimonial to the determination and cooperation of senior management, engineers and operators of all companies on the HiEnd critical path that DEC will come to market with their computer in the near future.

CHAPTER 6: CONCLUSIONS

1. M. Tyre. "Managing Innovation in the Manufacturing Environment: Creating Forums for Change on the Factory Floor," M.I.T. Sloan School of Management working paper #3005-89-BPS, December, 1989.

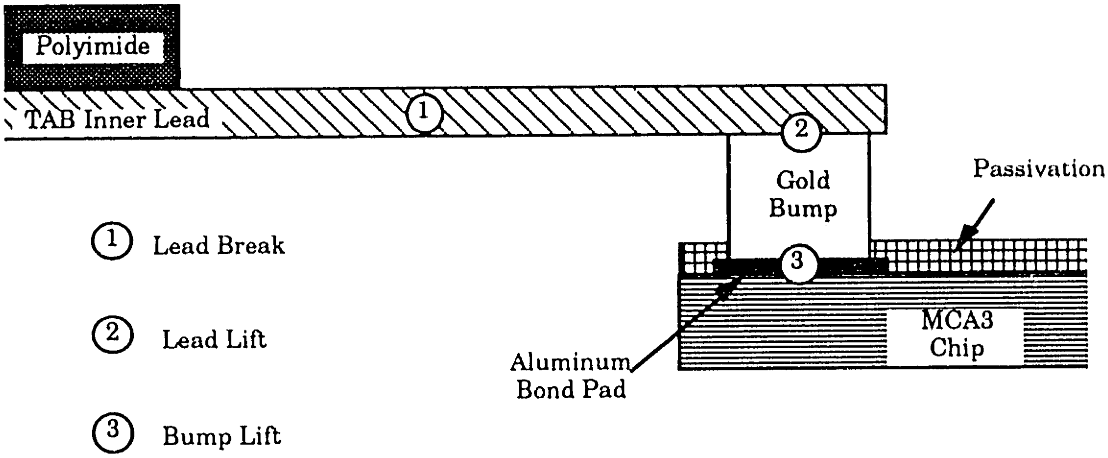
APPENDIX A: LEAD PULL INSPECTION

The lead pull test is a destructive measure of the integrity of an inner lead bond. The lead pull machine measures the force necessary to break a bonded lead. A small hook plucks a the lead from its bond pad. The force necessary to break the lead is recorded as data in an Excel© spreadsheet.

There are three common failure mechanisms in a lead pull test: lead breaks, lead lifts and bump lifts. Figure A.1 shows where these failures occur on a bonded lead. A lead break occurs along the length of the inner lead between the polyimide of the tape and the gold bump on the die. Lead breaks are generally considered to be "good" failures; they indicate that the bond strength was higher than the tensile strength of the lead. The lead lift occurs when the lead de-bonds from the gold bump. Lead lifts are generally indications of poor bond integrity, depending on the pull force at which they lift. The third common failure mode is the bump lift. Bump lifts occur when excessive force or temperature has been applied during inner lead bonding. The pressure cracks the passivation and/or the aluminum bond pad so that when the lead is pulled, the entire bump separates from the die. Similarly, excessive temperature may thermally shock the bond pads, also cracking them. Bump lifts occur most frequently when the temperature and pressure of the bonding operation must be adjusted up to minimize lead lifts.

APPENDIX A: LEAD PULL INSPECTION

Figure A.1: Lead Pull Failure Modes



A systematic error is associated with the numeric data generated in lead pull tests. The force necessary to pull the lead depends on the position of the lead pull hook along the length of the lead. The force the bonded region "sees" depends on the magnitude of the moment arm generated which, in turn, depends on the position of the hook between the bond region and the mylar tape.

One potential use of the lead pull data is as a process control metric. For this purpose, lead lifts are the only poor failure mechanism considered (bump lifts are also undesirable - see Table A.2). The pull force associated with a lead lift is weighted with a "WALD" number. A WALD Chart is a plot of the cumulative WALD index vs. some time parameter (possible parameters are the number of bonds performed by a thermode, lot numbers, date, etc.). The steepness of the resulting line reflects the degree of process control present - a very steep line indicates that the process is out of control. Motorola had not yet instituted process control using lead lift WALD by the time the LFM Program became involved in the Hyperchip project.

Motorola does make use of the lead pull data on a regular basis as a gate for incoming TAB tape batches. Lots are screened for the number of "undesirable" pull failures (lead lifts and bump lifts) and for the magnitude of their pull force.

Both numeric and attribute lead pull data were used extensively in this document to correlate TAB fabrication process parameters with bonding yield at Motorola. Chapter 2 and Chapter 3 discuss this effort in detail.