

**UNCLASSIFIED**

**AD 414397**

**DEFENSE DOCUMENTATION CENTER**

**FOR**

**SCIENTIFIC AND TECHNICAL INFORMATION**

**CAMERON STATION, ALEXANDRIA, VIRGINIA**



**UNCLASSIFIED**

**NOTICE:** When government or other drawings, specifications or other data are used for any purpose other than in connection with a definitely related government procurement operation, the U. S. Government thereby incurs no responsibility, nor any obligation whatsoever; and the fact that the Government may have formulated, furnished, or in any way supplied the said drawings, specifications, or other data is not to be regarded by implication or otherwise as in any manner licensing the holder or any other person or corporation, or conveying any rights or permission to manufacture, use or sell any patented invention that may in any way be related thereto.

63-4-5

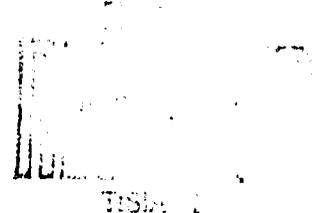
ADC-TDR-63-303

10 NE 1963

CATALOGED BY DDC

As ADJ No. \_\_\_\_\_

414397



FINAL REPORT ON

# Chinese-English Machine Translation Research

INTERNATIONAL BUSINESS MACHINES CORPORATION  
THOMAS J. WATSON RESEARCH CENTER, YORKTOWN HEIGHTS, NEW YORK

CONTRACT AF 30 (602)-2479

PREPARED FOR INFORMATION PROCESSING LABORATORY/RESEARCH AND TECHNOLOGY DIVISION  
ROME AIR DEVELOPMENT CENTER (AFSC)  
GRIFFISS AIR FORCE BASE, NEW YORK

IBM RESEARCH

sed  
n.

**RADC-TDR-63-303**

**June 10, 1963**

**Final Report on Chinese-English Machine Translation Research**

**International Business Machines Corporation  
Thomas J. Watson Research Center  
Yorktown Heights, New York**

**Contract AF 30(602)-2479**

**Prepared for**

**Information Processing Laboratory  
Research and Technology Division  
Rome Air Development Center (AFSC)**

**Griffiss Air Force Base  
New York**

RADC-TDR-63-303

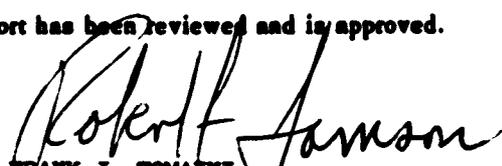
## ABSTRACT

A study of the requirements of an operational Chinese-English machine translation system utilizing the techniques of the AN/GSQ-16 (XW 2) Language Processor has led to the design and fabrication of a Chinese Character Coding Device, the Sinowriter, and the compilation of a Chinese-English multipass lexicon based on grammatical analysis. The results of the input coding study are presented, along with linguistic research studies and multipass Chinese-English processor techniques. A section of this report will be devoted to recommendations for future Sinowriter systems and for linguistic as well as lexicographical efforts required to ultimately perform Chinese-English machine translation on a random basis.

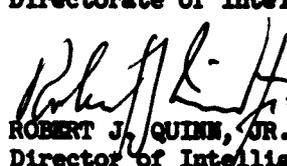
## PUBLICATION REVIEW

This report has been reviewed and is approved.

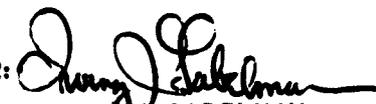
Approved:

  
 FRANK J. TOMAINI  
 Chief, Information Processing Laboratory  
 Directorate of Intelligence & Electronic Warfare

Approved:

  
 ROBERT J. QUINN, JR., Col, USAF  
 Director of Intelligence & Electronic Warfare

FOR THE COMMANDER:

  
 IRVING J. GABELMAN  
 Director of Advanced Studies

## FOREWORD

This report is intended to document the state of development of the Chinese-English Machine Translation research program at the termination of Contract AF 30(602)-2479. The objectives of the Contract are twofold:

1. Development and fabrication of a Chinese Character Coding Device for the preparation of Chinese texts into machine readable forms, and,
2. Study and compilation of a Chinese-English multipass lexicon based on grammatical analysis.

The work has been carried on in two successive phases covering a period of two years. In the first year of the Contract, a Chinese Character Coding Device, the Sinowriter, was developed jointly by IBM Research and the Mergenthaler Linotype Company of Brooklyn, New York. An experimental model was delivered to the Air Force on April 11, 1963. Linguistic research and multipass lexicon work were initiated at the same time but did not really pick up momentum until the beginning of the second Contract year. This work resulted in two lexicons, the one compiled during the second Contract year represents the beginning of an operational Chinese-English machine translation system for the random translation of Chinese. Both lexicons are based on grammatical analysis and both are based on the analysis of limited input texts. The experimental lexicon contains about 1400 entries and the more extensive lexicon about 10,000 entries.

The Chinese character-indexing method used for the Sinowriter is an adaption of a modified Ming Kwai Chinese typewriter scheme proposed by Dr. Lin Yutang.

The multipass lexicon analysis techniques are based on those of the AN/GSQ-16 (XW 2) Language Processor System. The linguistic analysis routines are divided into the following passes: 1. The discontinuous constituent and parts of speech pass, 2. The auxiliary word pass, 3. The functive 了 (le) pass, 4. The functive 的 (de) pass, 5. English synthesis pass.

The first phase of the development work, the development and fabrication of a Chinese Character Coding Device -- the Sinowriter -- was jointly performed by IBM Research and the Mergenthaler Linotype Company. The second phase of the Contract work was conducted by IBM Research personnel under the direction of Dr. Ernest H. Goldman, director of the Experimental Systems department of IBM Research. Appendix III lists all personnel associated with the project.

## TABLE OF CONTENTS

	<b>Page</b>
<b>I. Introduction</b>	<b>1</b>
1.1 Input Coding Aspects	1
1.2 Linguistic Aspects	2
<b>II. Chinese Character Coding Device -- The Sinowriter</b>	<b>3</b>
2.1 Sinowriter Chinese Character Selection Scheme	3
2.2 Sinowriter Input Device Chinese Character Vocabulary	9
2.3 Sinowriter Keyboard Design and Binary Coding Assignments	12
2.4 Evaluative Study of the Sinowriter Keyboard Chinese Character Indexing Method	15
<b>III. The Sinowriter -- Engineering and Fabrication</b>	<b>17</b>
3.1 Design Requirements	17
3.2 Design Goals	18
3.3 Execution	20
3.4 Mechanical Requirements	30
3.5 Coding Arrangement for Word Acquisition	40
3.6 Problems	49
3.7 Quality Control	57
<b>IV. Classification of Characters</b>	<b>61</b>
4.1 Classification by Radicals	61
4.2 Classification by Stroke	62
4.3 Classification by Graphical Recognition	63
4.4 Classification by Number	64

	<b>Page</b>
V. Estimate of the Total Number of Characters Employed in Chinese Scientific Texts	65
VI. Lexical Processing	67
6.1 Preparation of Input Material	67
6.2 Detection of Input Errors by Means of a 704 Unidirectional Single-Pass Translation	67
6.3 Master Reference File	67
6.4 Generation of Multipass Entries from the Master Reference File	71
VII. Linguistic Processing	75
7.1 Parts-of-Speech	75
7.2 The XVN and DC Pass	78
7.3 The 的 Pass	80
7.4 English Synthesis	83
VIII. Recommendations and Conclusions	87
8.1 Software Considerations	87
8.2 Hardware Considerations	92
Appendix I. - The Sinowriter (Illustrations)	95
Appendix II. - Samples of Chinese-English Machine Translation Output	113
Appendix III. - Project Personnel	129

## List of Illustrations

Figure 2.1	Character Top and Bottom Characteristics in Quadrants	4
Figure 2.2	Character Top and Bottom Configurations According to Groups	6
Figure 2.3	Sinowriter Keyboard Design by Zones	7
Figure 2.4	Example of a Sinowriter Key	8
Figure 2.5	Examples of Characters	13
Figure 3.1	Chinese Character Code Group Format as Punched on Paper Tape	19
Figure 3.2	Number of Characters Per Family	21
Figure 3.3	Diagram of Chinese Input Device	36
Figure 3.4	Address Codes Photographic Memory Plates	41
Figure 3.5	Relative Logical Space Available	43
Figure 3.6	Overlay of Areas Depicted in Figure 3.5	44
Figure 3.7	Diagrammatic Representation of Coding Problem Area	46
Figure 3.8	Diagrammatic Representation of Coding Problem Area	47
Figure 3.9	Roof Mirror	53
Figure 6.1	Error Detection by Unidirectional Single-Pass Translation	68
Figure 6.2	Multipass Simulation by 704 Computer	70
Appendix I	Photographs of Sinowriter	95

## List of Tables

Table 2.1	Typical Patterns of Upper Key Characteristics	10
Table 2.2	Typical Patterns of Lower Key Characteristics	11
Table 3.1	1, 2, 4, 8, 10 Binary	26
Table 3.2	1, 2, 4, 8, 8 Binary	27
Table 5.1	List of Special Term Dictionaries	66

## I. INTRODUCTION

To develop the ability to translate Chinese by machine it is necessary to carry out three major tasks. The first is to develop the physical means for converting an indexing scheme for Chinese characters into binary codes. The second is to investigate and overcome the syntactic problems of the Chinese language. The third is to construct an appropriate Chinese-English Machine Translation dictionary to render translation on a random basis.

### 1.1 Input Coding Aspects

To design an efficient Chinese coding device, the following criteria should be borne in mind:

1. Minimal keyboard ambiguities
2. Speed and ease of character selection
3. Flexibility in handling both conventional and simplified characters
4. Minimal learning time required for operator training
5. Ability for use by operators of any nationality
6. Simplicity of the keyboard design, approaching that of an alphabetic language typewriter.

To achieve these goals and also to design an input device the output of which will be compatible with the AN/GSQ-16 (XW-2) Language Processor system, IBM Research in cooperation with the Mergenthaler Linotype Company has selected the character indexing scheme proposed by Dr. Lin Yutang. The Mergenthaler Linotype Company has built the Sinowriter, a Chinese character coding device.

The indexing scheme utilized relies on geometric character recognition. Each character is grouped into any one of the one thousand odd families, according to character top, bottom, and numeral configurations.

The Sinowriter input device has a vocabulary of approximately 6500 characters which is adequate for most modern scientific and everyday common Chinese publications.

## 1.2 Linguistic Aspects

At the present stage of machine translation research it is obvious that dictionary construction cannot be divorced from syntactic considerations. For this reason, great emphasis has been put, in the lexicon compilation phase of the research work, on the study of grammatical problems encountered in modern Chinese text. Initially, the philosophy has been to attack these machine translation problems on a modest but continually expanding basis. Thus the work has been oriented toward a continually expanding corpus of Chinese texts.

Specific problems connected with Chinese grammar that have been attacked are categorized into the following classes:

1. Discontinuous constituents
2. Parts of speech ambiguities
3. Auxiliary word problems
4. Functive words
5. English synthesis studies

A tentative parts of speech classification for machine translation purposes has been decided upon and the results have been incorporated into the multipass Chinese-English lexicon.

Although all the analysis routines and dictionary entries designed are based on the techniques of the AN/GSQ-16 (XW-2) Language Processor, the existence of a Master Reference File makes the conversion of these entries to that of any other system a relatively simple task.

## II. CHINESE CHARACTER CODING DEVICE -- THE SINOWRITER

The Sinowriter is a device for converting Chinese Character texts into machine codes as inputs to the AN/GSQ-16 (XW-2) Language Processor. The description of the device can be divided into two main parts: the Chinese character classification scheme and the Chinese character vocabulary; the optical, mechanical, and electrical systems of the device.

### 2.1 Sinowriter Chinese Character Selection Scheme

The basic Sinowriter Chinese character selection scheme makes use of a modified version of the Lin Yutang top-bottom geometric recognition method. The Chinese character vocabulary of the device totals about 6500 conventional and simplified characters. These are grouped into about 1000 character families according to 36 top characteristic configurations of the characters. Within each family all characters have the same top and bottom characteristics and in most cases the number of characters in a family is under sixteen. Rarely is it over sixteen. There are about eighteen families with about thirty-two characters each. In each family the characters are numbered from one to sixteen with the most frequently used characters occupying the lower numbers.

The basic Chinese character indexing scheme used in the Sinowriter input device is a member of the geometric character recognition family. The other popular member of this family is the Wang, Yun-Wu Four Corner indexing method which classifies Chinese characters according to their respective four corners. A description and evaluation of the Four Corner method and various other methods will be given in a later section. The Sinowriter indexing scheme considers much more than character corners: it looks instead at the whole character in relation to a unit square area divided into

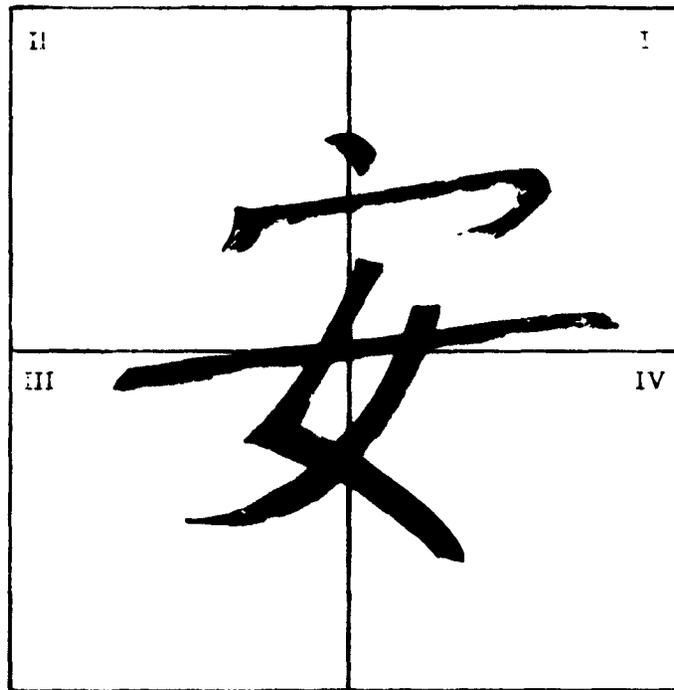


Figure 2.1 Character Top and Bottom Characteristics in Quadrants

quadrants, i. e., 1st, 2nd, 3rd, and 4th. The top and bottom character configurations or characteristics are selected by an examination of the character geometric patterns contained therein.

An illustration of a character unit square and its component quadrants is given in Figure 2.1 shown on opposite page.

The top characteristics of a character can usually be found in quadrants (I) and (II), in (II) and (III) or in (II) alone. The bottom characteristics can usually be found in quadrants (III) and (IV), in (IV) and (I), or in (IV) alone. These are not to be regarded as definitive rules but rather as a set of guidelines for quick recognition of the characteristics of the characters in question. The thirty-six top characteristics can be further classified for easy identification into five subgroups: the horizontal stroke group —, the vertical stroke group |, the slanting stroke group /, the dot stroke group ·, and the radical group. The bottom characteristics are similarly grouped according to strokes. Figure 2.2 shows the respective top and bottom character stroke patterns and the groups they make up. In Figure 2.3 is shown a Sinowriter keyboard design separated by dotted lines into five key stroke zones based on the top characteristics. Zone I comprises twelve keys whose top or bottom characteristics are frequently used radicals. Zone II includes six keys whose top characteristics have dots in common. Zone III has all horizontal stroke keys. Zone IV contains slanting strokes. And finally Zone V contains all with straight line characteristic tops. As can be seen in Figure 2.3 each key tap is further subdivided into three areas, the thirty-six top characteristics are to the left of the diagonal lines on the key surface, the thirty bottom characteristics are to the right of the diagonal lines, and the numeral characteristics, from 1 to 36, appear in the squares on the upper right-hand corners of the keys.

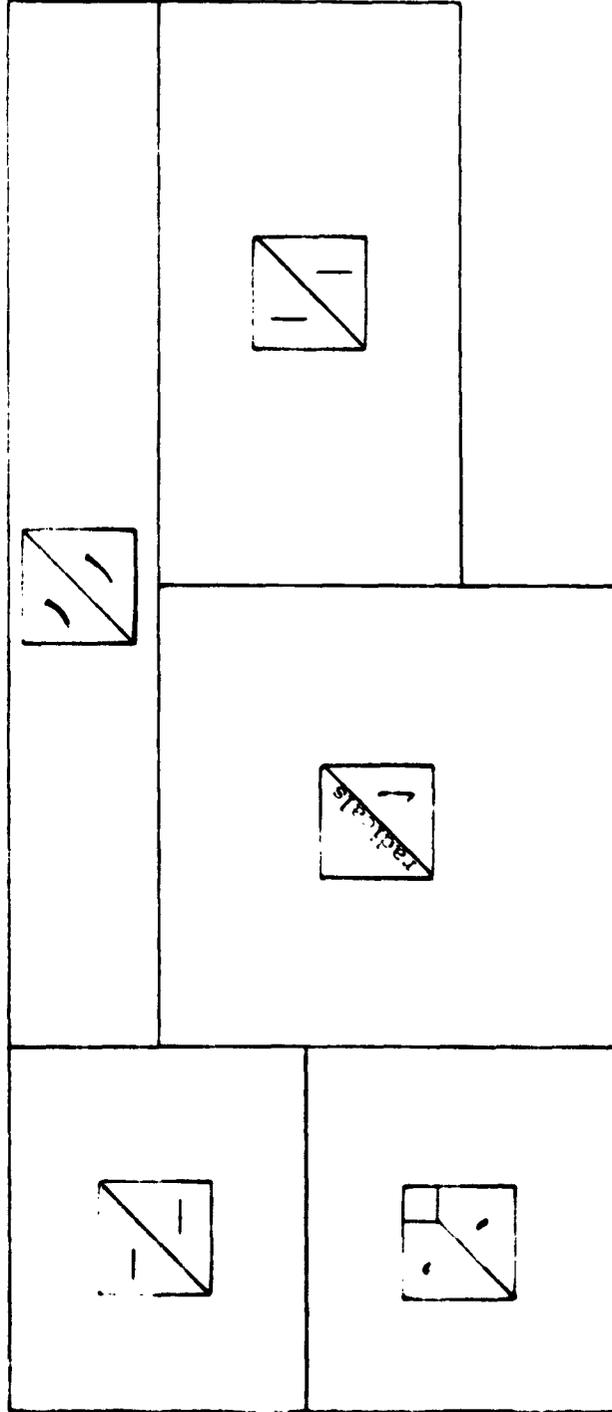


Figure 2.2 Character Top and Bottom Configurations According to Groups



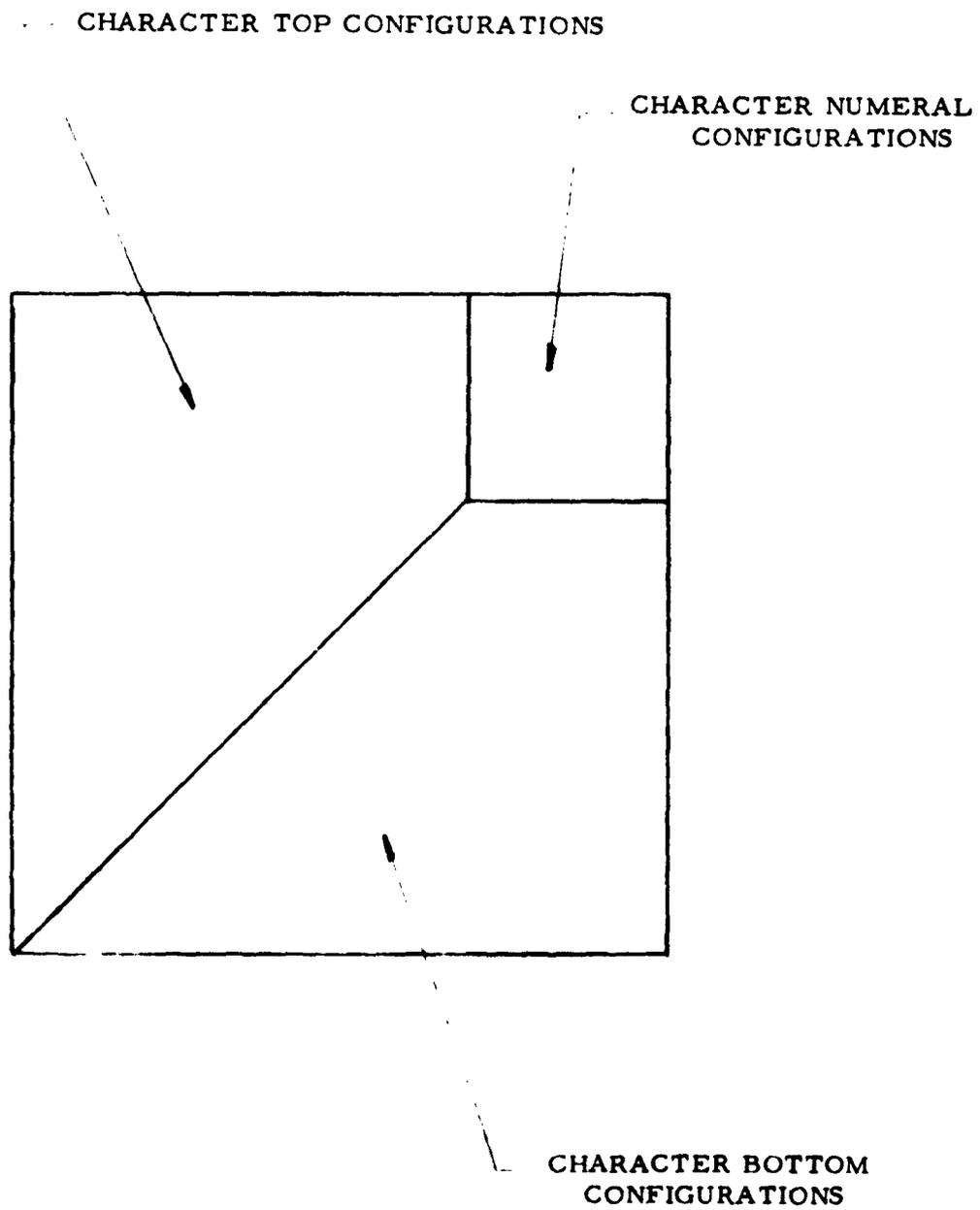


Figure 2.4 Example of a Sinewriter Key

Figure 2. 4 is a typical Sinowriter key with the subareas indicated to show their respective, intended purposes.

The following examples illustrate the method of selection of Chinese characters by means of the Sinowriter indexing scheme and also the operations of the Sinowriter keyboard.

1. 點; 回, 口, 2
2. 式; 乚, 乚, 6
3. 相; 木, 口, 1
4. 立; 一, 一, 1
5. 之; 丶, 一, 3

Other examples of the use of the top and bottom characteristics are shown in Tables 2. 1 and 2. 2.

## 2. 2 Sinowriter Input Device Chinese Character Vocabulary

As was pointed out previously, the Sinowriter input device has a vocabulary of about 6500 characters, grouped into a total of about 1000 families. The distribution of characters by families of various sizes (from one character to thirty-two characters) is shown in Figure 3. 2. In the case of families of more than sixteen characters, the selection of characters therein is achieved by depressing the second page button on the console. The Sinowriter character population also includes 528 simplified characters and radicals. These are necessary because the Communist Chinese, in their mass education movement, have re-introduced and popularized the use of the simplified writing of characters, and these simplified characters are employed extensively in today's mainland China publications.

Table 2.1 Typical Patterns of Upper Key Character

<p>(6) <b>イ</b></p> <p>イ 仁</p>	<p>(5) <b>宀</b></p> <p>宀 宀 宀 买 軍 密</p>	<p>(4) <b>雲</b></p> <p>雲 雲</p>	<p>(3) <b>田</b></p> <p>田 田 田 罰 野 問</p>	<p>(2) <b>貝</b></p> <p>貝 財</p>
<p>(12) <b>木</b></p> <p>木 木 松 札</p>	<p>(11) <b>才</b></p> <p>才 把</p>	<p>(10) <b>艸</b></p> <p>艸 艸 蓀 茶 范 薪 花 若 薈 芳</p>	<p>(9) <b>艸</b></p> <p>艸 艸 萌 荳 華 蕙</p>	<p>(8) <b>艸</b></p>
<p>(18) <b>一</b></p> <p>一 一 一 一 一 頭 元 不 碼 珍 豕 頭 云 更 破 列 豕</p>	<p>(17) <b>几</b></p> <p>几 几 几 几 几 鳳 丞 盈 階 孫 凡 鳳 登 朵 陳 孔 乙</p>	<p>(16) <b>冫</b></p> <p>冫 冫 冫 冫 冫 冫 刑 配 要 功 五 耶 現 刑 刑 刑 刑 刑 刑</p>	<p>(15) <b>厂</b></p> <p>厂 厂 厂 厂 厂 歷 馳 匪 卧 髮 原 驛 醫 腎 鬚</p>	<p>(14) <b>尸</b></p> <p>尸 尸 尸 尸 尸 巽 尼 彈 引 弱</p>
<p>(24) <b>艸</b></p> <p>艸 艸 艸 艸 艸 感 中 蟲 薰 精 羊 感 中 中 少 迷 判</p>	<p>(23) <b>山</b></p> <p>山 山 山 山 斷 將 豈 幽 斷 裝 崎 凹</p>	<p>(22) <b>艸</b></p> <p>艸 艸 艸 艸 艸 順 革 甘 業 悲 弗 州 州 州 州 州</p>	<p>(21) <b>牛</b></p> <p>牛 牛 牛 牛 牛 牲 章 虎 鹵 齒 上 靠 建 虛 虛 齡 皆</p>	<p>(20) <b>車</b></p> <p>車 車 車 車 車 耨 耨 耨 耨 耨 耨 耨 耨 耨 耨</p>
<p>(30) <b>彳</b></p> <p>彳 彳 彳 彳 彳 行 狗 必 身 舟 舟 舟 舟 射 血 船 帛 師 師 師 師</p>	<p>(29) <b>彳</b></p> <p>彳 彳 彳 彳 彳 終 能 婦 巢 網 迨 妨 笛</p>	<p>(28) <b>彳</b></p> <p>彳 彳 彳 彳 彳 彳 彳 缺 年 氣 乞 魚 角 夕 夕 夕 夕 鮮 解 怨 旬 息 祭 久 免</p>	<p>(27) <b>力</b></p> <p>力 力 力 力 力 力 力 大 大 大 大 大 大 大 奈 在 春 祭 勞 火 片 片 查 尤 奉 脊 營 炒 版 央</p>	<p>(26) <b>干</b></p> <p>干 干 干 干 干 干 禾 禾 禾 禾 重 香 香 香 香</p>
<p>(26) <b>氵</b></p> <p>氵 氵 氵 河 永 良 漁</p>	<p>(35) <b>心</b></p> <p>心 心 心 心 心 心 心 譽 義 馮 凶 盆 釜 學 並 姿 川 翁 谷</p>	<p>(34) <b>才</b></p> <p>才 才 才 才 才 才 才 祇 視 齊 音 款 房 旗 禮 觀 坡 葬 產 產 放</p>	<p>(33) <b>十</b></p> <p>十 十 十 十 十 十 十 亡 市 言 言 言 言 言 望 哀 高 語 遠 遠 遠</p>	<p>(3) <b>十</b></p>



Table 2.2 Typical Patterns of Lower Key Character

<p>(6) 儿 儿</p> <p>儿 儿 魁 兄 魁 兆</p>	<p>(5) 丷 丷</p> <p>丷 匕 丷 風 瓦 此 飽 鼠 尤</p>	<p>(4) 司 司 羽</p> <p>羽 司 母 勿 司 羽 騎 伺 悔 傷 局 仍 悔 笏 句 弓</p>	<p>(3) 冂 冂</p> <p>冂 冂 冂 冂 用 冂 寓 桶 篇 萬 高</p>
<p>(12) 木 木</p> <p>木 木 除 栗 蝶 柔</p>	<p>(11) 寸 寸</p> <p>寸 寺 將</p>	<p>(10) 可 子 丿 丿</p> <p>可 子 丿 丿 倚 字 刪 打 何 厚 剛 牙</p>	<p>(9) 乚 乚</p> <p>乚 乚 代 戰 民 賤</p>
<p>(18) 一 一</p> <p>一 一 一 一 丑 盡 虹 些 上 血 空 畫</p>	<p>(17) 山 山 山</p> <p>山 山 山 巨 踏 掘 凹 世 閭 岳 屈</p>	<p>(16) 土 土 土</p> <p>佳 主 王 土 士 催 星 皇 吐 垂 離 僅</p>	<p>(15) 口 口</p> <p>口 回 君</p>
<p>(24) 女 丿 山 山</p> <p>女 丿 虫 山 接 抄 強 云 因 彬 騷 弦</p>	<p>(23) 冫 冫</p> <p>冫 冫 冰 今</p>	<p>(22) 又 又 又 又 又</p> <p>良 及 文 又 又 服 援 敢 校 發 慢 板 散 駁 慢</p>	<p>(21) 斤 巾 巾 巾 川 川 川</p> <p>斤 巾 巾 巾 川 川 川 折 即 掃 卧 馴 介 引 折 命 掛 非</p>
<p>(30) 小 小 小</p> <p>小 小 不 係 來 杯 你 東 坏</p>	<p>(29) 水 水 冫 冫 冫</p> <p>水 水 冫 冫 冫 錄 樣 家 眾 衣 食 漆 水 啄 聚 嚷 長</p>	<p>(28) 冫 冫 冫</p> <p>冫 冫 冫 冬 魚 專 包 燕 於</p>	<p>(27) 冫 冫 冫 冫</p> <p>冫 冫 冫 冫 孤 乏 道 建 題 爪</p>

2.2 Typical Patterns of Lower Key Characteristics

<p>(4)</p> <p>羽 司 母 勿 可 了</p> <p>翳 伺 悔 傷 局 仍</p> <p>翳 伺 悔 傷 局 仍</p>	<p>(3)</p> <p>冏 用 冊 冏</p> <p>冏 用 冊 冏</p> <p>冏 用 冊 冏</p>	<p>(2)</p> <p>月 月</p> <p>月 月</p> <p>月 月</p>	<p>(1)</p> <p>冏 冏</p> <p>馬 冏 鳥</p> <p>媽 馬 鸚</p>	<p>LOWER KEY</p> <p>TYPICAL PATTERN</p> <p>EXAMPLES</p>
<p>(10)</p> <p>可 子 冫 冫</p> <p>倚 字 刪 打</p> <p>倚 字 刪 打</p>	<p>(9)</p> <p>匕 戈</p> <p>代 戰</p> <p>代 戰</p>	<p>(8)</p> <p>心 心</p> <p>必 心</p> <p>秘 思</p> <p>秘 思</p>	<p>(7)</p> <p>力 力 力</p> <p>力 刀 力</p> <p>芳 芬 功</p> <p>芳 芬 功</p>	<p>LOWER KEY</p> <p>TYPICAL PATTERN</p> <p>EXAMPLES</p>
<p>(16)</p> <p>佳 主 王 土 士</p> <p>催 星 皇 吐 垂</p> <p>催 星 皇 吐 垂</p>	<p>(15)</p> <p>口</p> <p>口</p> <p>回 君</p>	<p>(14)</p> <p>冫</p> <p>冫</p> <p>当</p>	<p>(13)</p> <p>冫 田 日 日 冫</p> <p>腦 富 眉 昔 曲</p> <p>腦 富 眉 昔 曲</p>	<p>LOWER KEY</p> <p>TYPICAL PATTERN</p> <p>EXAMPLES</p>
<p>(22)</p> <p>冫 及 文 又 又</p> <p>服 援 敢 校 發</p> <p>服 援 敢 校 發</p>	<p>(21)</p> <p>斤 巾 巾 巾 川 川 川 冫</p> <p>折 即 掃 卧 馴 介 引</p> <p>折 即 掃 卧 馴 介 引</p>	<p>(20)</p> <p>十 十 十 十</p> <p>擇 伴 昇 軍</p> <p>擇 伴 昇 軍</p>	<p>(19)</p> <p>冫</p> <p>冫</p> <p>那</p>	<p>LOWER KEY</p> <p>TYPICAL PATTERN</p> <p>EXAMPLES</p>
<p>(28)</p> <p>冫 冫 冫</p> <p>年 冬 魚 專</p> <p>年 冬 魚 專</p>	<p>(27)</p> <p>冫 冫 冫 冫 冫</p> <p>孤 乏 道 建 題</p> <p>孤 乏 道 建 題</p>	<p>(26)</p> <p>冫 冫</p> <p>扒 項</p> <p>扒 項</p>	<p>(25)</p> <p>冫 冫</p> <p>咳 吹</p> <p>咳 吹</p>	<p>LOWER KEY</p> <p>TYPICAL PATTERN</p> <p>EXAMPLES</p>

2

It is, therefore, obvious that Chinese character input devices not only should possess the flexibility of processing both conventional and simplified characters but should also include in the vocabulary those character radicals most frequently used in simplified form. Some examples of characters in their conventional as well as simplified form are shown in Figure 2.5.

The simplified forms of characters are being generated almost daily and it is obvious that the input character vocabulary should be updated at regular intervals to process the most recent publications.

### 2.3 Sinowriter Keyboard Design and Binary Coding Assignments

In the assignment of binary codes to the various top, bottom and numeral characteristics, conventional Flexowriter codes have been employed.

The terminal-type punctuation marks have been reserved strictly for punctuation purposes, in order to prevent confusion between these and a Chinese character code. Because of the redundant nature of the codes assigned to the various top, bottom, and numeral characteristics, it is possible for a Chinese character to be represented by a string of three identical codes, i. e., a character's top, bottom, and numeral characteristics may happen to fall on the same key.

The Sinowriter keyboard is designed to operate in two modes, either the Chinese mode or the English mode. All input English characters, Arabic numerals as well as punctuation marks, are considered as English inputs. The English mode of operation has to be entered by depressing the English Mode console key. In this mode the keyboard functions like a regular typewriter.

CHARACTERS	CONVENTIONAL	SIMPLIFIED
導	導	导
廠	廠	厂
華	華	华
龍	龍	龙
書	書	书
廳	廳	厅
農	農	农
從	從	从

Figure 2.5 Examples of Characters

After the Chinese Mode button on the console has been depressed, the circuit switches to a three-stage buffer register and the keyboard operations are altered slightly. The Flexowriter will withhold punching a code in the paper tape until three keys have been depressed in succession. Three codes corresponding to the keys depressed will be punched and in addition a "space" code will automatically be punched. The Sinowriter keyboard will stay in the Chinese mode of operation until the English Mode key is depressed.

Since some of the punctuation code keys have to be shared as Sinocode code keys, the distinction of these key codes when used as Sinocodes or punctuation codes will have to be made clear and unique. For Chinese input in the AN/GSQ-16 (XW-2) system the various punctuation marks are designated as follows:

<u>Punctuation Marks</u>	<u>AN/GSQ-16 (XW-2)</u>
(,)	(LC)(,)
(.)	(LC)(.)
(space)	(LC)(space)
(	(LC)(UC)(LC)9
)	(LC)(UC)(LC)(Z)
(')	(LC)(UC)(LC)2
''	(LC)(UC)(LC)22
(*)	(LC)(UC)(LC)8
(-)	(LC)(-)
(:)	(LC)(UC)(LC)(:)
(;)	(LC)(;)
(/)	(LC)(/)

The terminal punctuations are prepared as follows:

<u>Terminal Punctuations</u>	<u>AN/GSQ-16 (XW-2)</u>
(.)(space)(space)(cap)	(LC)(.)(space)(space)(cap)
(.)(par)(tab)(cap)	(LC)(par)(tab)(cap)
(?)(space)(space)(cap)	(LC)(UC)(LC)(?)(space)(space)(cap)
(?)(par)(tab)(cap)	(LC)(UC)(LC)(?)(par)(tab)(cap)
(!)(space)(space)(cap)	(LC)(UC)(LC)(space)(space)(cap)
(!)(par)(tab)(cap)	(LC)(UC)(LC)(!)(par)(tab)(cap)

The beginning of sentence indications are of two types, (space)(space)(cap) or (par)(tab)(cap).

Figure 2.3 shows the Sinowriter keyboard design.

On each key a maximum of five fields are shown:

top characteristics, bottom characteristics, numeral characteristics, octal number of the six-bit code, and the corresponding English letter on a conventional keyboard.

#### 2.4 Evaluative Study of the Sinowriter Keyboard Chinese Character Indexing Method

The keyboard indexing scheme of the Sinowriter which has previously been described has been tested extensively by means of manual simulation. A total of about 5000 lexicon entries have been compiled using this method. These entries compound Chinese words made up of from one to seven characters. The total number of characters used to conduct this manual encodement of Chinese texts by the Sinowriter method is about 4 x 5000 entries. A number of interesting and previously hidden aspects of the scheme have been uncovered. A number of very encouraging conclusions were also discovered, namely:

1. The Sinowriter character vocabulary as it stands today should be modified extensively. Many of the rarely used characters which the Sinowriter vocabulary contains should be eliminated and substituted by modern up-to-date and more frequently used ones.
2. A reclassification will have to be made to eliminate some of the present ambiguities.
3. Some of the 20,000 characters that IBM encoded were not found in the vocabulary. Some of these are the simplified characters which have just recently come into popular use, the rest are conventional forms.
4. Mathematical signs, symbols, and the Greek alphabet should definitely be added.
5. Ways and means should be provided to care for the input characters which are not found by the Sinowriter.
6. The Sinowriter classification scheme is basically sound and efficient.
7. The keyboard can be learned and operated by personnel of almost any nationality.

In the Recommendation Section of this report, there is given a more detailed description of the problems and their recommended methods of solution.

### III. THE SINOWRITER - ENGINEERING AND FABRICATION

The Chinese Sinowriter was subcontracted to the Mergenthaler Linotype Company for design and fabrication. This section describes the engineering and operating characteristics of the input device as reported by the Mergenthaler Linotype Company.

#### 3.1 Design Requirements

##### 3.1.1 Language Classification

The "Ming Kwai" system of character classification designed by the Chinese scholar Lin Yutang forms the basis of the code system and memory configuration. This classification system is essentially based on the geometric properties of the written Chinese characters. In essence, the characters (Chinese words) are arranged in groups which exhibit geometric characteristics in the upper (or upper left) and lower (or lower right) areas. The original "Ming Kwai" keyboard used 36 upper characteristics and 28 lower characteristics, whereas the present design utilizes 36 upper characteristics and 30 lower characteristics. The resulting groups are referred to as families.

##### 3.1.2 Keyboard

The Friden Flexowriter is utilized in this design, because this machine is equipped with a paper punch mechanism and a coding system of great flexibility. The Flexowriter is equipped with a reader head which can be used to check system operation. In addition, the output of the keyboard coding system is easily manipulated to meet external system requirements.

## 3.2 Design Goals

### 3.2.1 Control System

The punched paper tape coding was to consist of 3 digits and a fixed code indicating word ending. Figure 3.1. Besides using the keyboard to code Chinese, it was to be used for the normal keyboarding of English.

The keyboard was to be used as the system control: pressing a prescribed key (upper case) would place the system in the "Chinese" code mode of operation, another key (lower case) in the "English" code mode. The control system was to operate as fast as the keyboard punch combination so that it would not limit the system speed of operation.

Additional controls were to be provided where necessary, (i. e., provide for plate two selection and error erasure plus a master power on and off switch. )

### 3.2.2 Optical System

The optical system was to be designed after the optical wedge scanning system patented by Mr. Edward Bechtold. This system was to be extended in order to accommodate additional characters in the x-y coordinate plane. The number of coordinate positions to be available was to be a minimum of one thousand and eighty plus a sufficient amount to permit additions to the basic vocabulary amounting to 20 per cent of the minimum number. A capability for expansion to approximately thirty thousand characters was to be shown.

The most complicated character would probably not exceed 40 strokes, it is assumed that half of these strokes will be in the

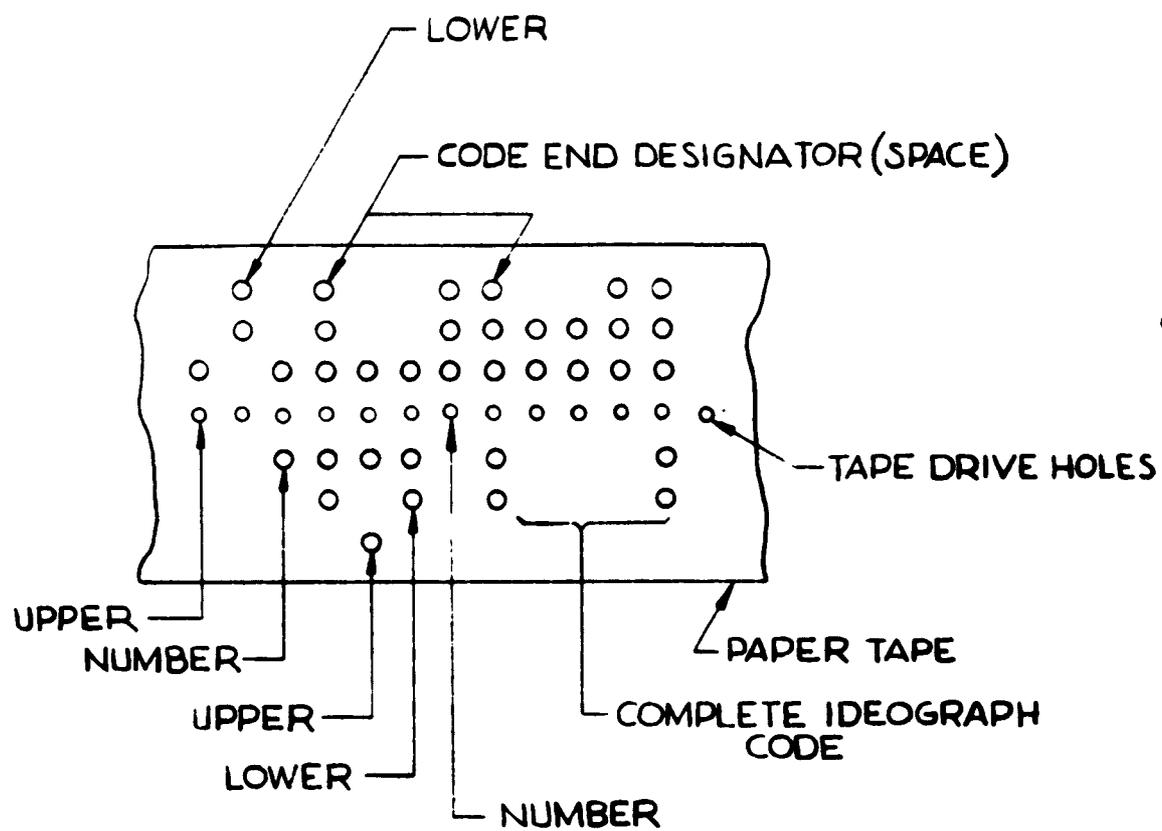


Figure 3.1 Chinese Character Code Group Format as Punched on Paper Tape

vertical and half in the horizontal axes of the character. Character size would not be smaller than  $3/64$  (approximately 1/mm). The resultant required resolution is approximately 20 lines/mm. Therefore the optical system resolution should not be required to exceed 20 lines/mm at the focal plane at the objective.

### 3.2.3 The Mechanism

The mechanism was to be as simple as possible consistent with the optical and control system requirement in view of the basic premise that the delivered hardware would be considered as elementary breadboard construction.

## 3.3 Execution

### 3.3.1 Vocabulary Preparation

The Lin Yutang "Ming Kwai" vocabulary had been classified for use in an earlier mechanical version of a Chinese typewriter and was immediately available. This vocabulary was studied and reclassified in two ways based on the studies of the practical use of the keyboard. Sources of error in application of the keyboard were: confusion as to actual classification symbol key to be applied, and inexact classification of the characters due to the necessity of mixing of extensive families, caused by insufficient accommodation in family group space allocations in the typewriter mechanism.

The basic storage medium in this system is a photographic plate on which all characters comprising the vocabulary are placed. Therefore the family size would be restricted only by the optical field available at any coordinate designation. Since the "Ming Kwai" typewriter's vocabulary family capability of eight characters per

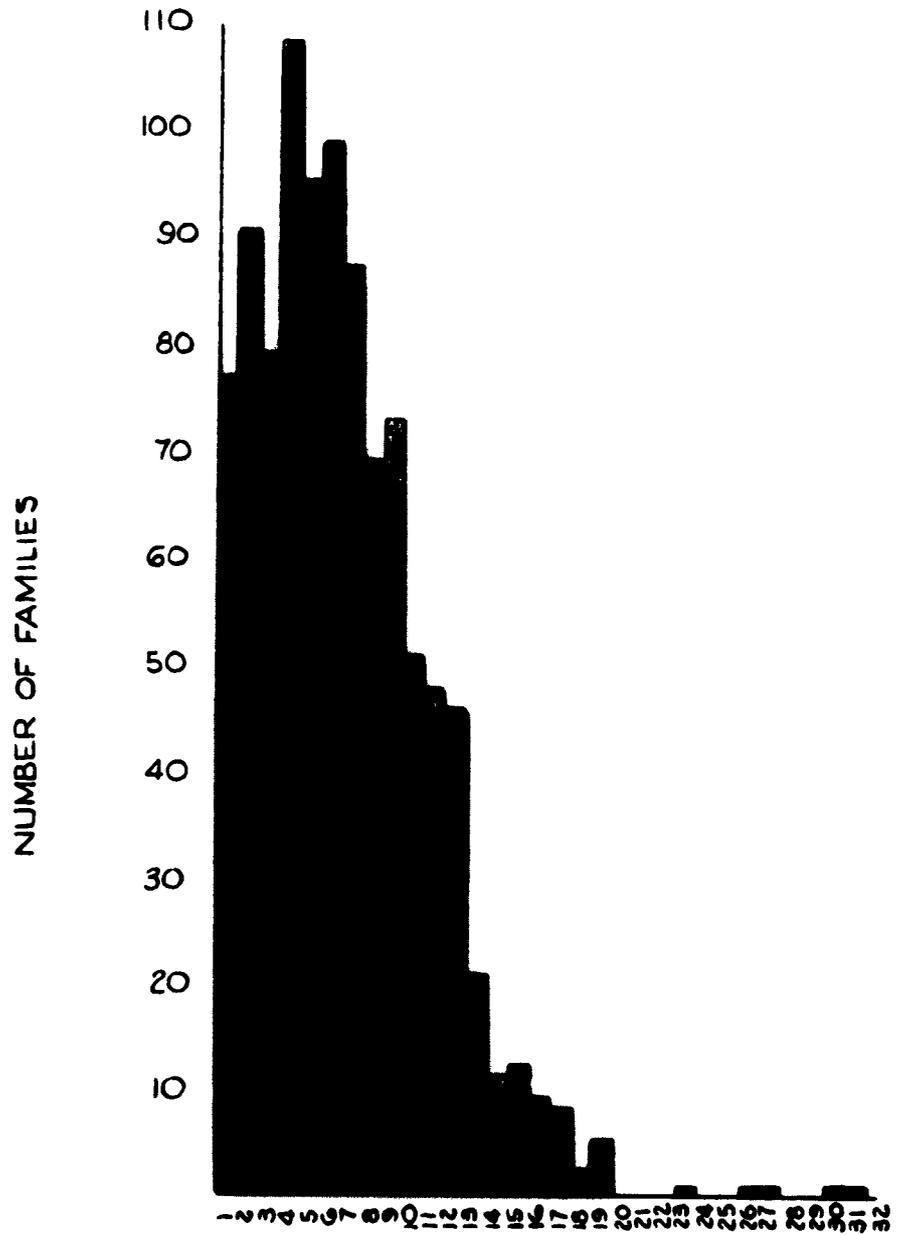


Figure 3.2 Number of Characters Per Family

family proved inadequate, it was decided that a maximum family size was to be a four by four matrix resulting in 16 characters per family. It was felt (and verified) that very few family groups would consist of more than 16 characters.

By extending the size of the family groups to 16, confusion due to "spill over" is avoided or reduced to a minimum. Extending the family sizes in addition permits us to eliminate the errors resulting from use of an ambiguous symbol for the word configuration. Errors of this type required the operator of the "Ming Kwai" typewriter to try other symbols of approximately the same configuration until the desired character was located. A simple expedient to eliminate this problem is to assign "problem" characters two addresses thus making each of them a member of two families.

It is obvious that, by extending family sizes and accepting two code addresses as valid for a single ideograph, the major problems encountered in the "Ming Kwai" typewriter are circumvented.

To include 540 simplified characters not within the original vocabulary, a slight alteration in the number of symbols applied to the keyboard was made. The upper characteristics remained 36 while the lower was expanded from 28 to 30.

A dictionary was prepared using master sheets, showing word space allocations in the family group, in concert with the upper and lower characteristic designations. In essence, the individual ideograph address is determined at this point in the work.

A master family pasteup was now made in accordance with the dictionary, with each character approximately 1" by 1". The family master occupied a square area 4" by 4". Each master was

photographed on 4 to 1 reduction resulting in a secondary master 1 inch square. In turn these 1 inch square masters were pasted up on an aluminum plate in the required x-y coordinate system determined by the scanner optics.

The assembled master vocabulary plates were in turn photographed on a scale 5.03 to 1. The resulting photographic memory plate size therefore was  $6 \frac{11}{16}$ " x  $6 \frac{3}{4}$ " across the major field. Individual characters are approximately .049 inch square.

### 3.3.2 Mechanism

**Optical System Requirements:** The optics were to be designed to resolve 20 lines/mm at the focal plane. This would not require excessive resolution from the photographic reproduction processes. The basic family size should be such that an individual character would not be reduced to a size less than .040 inch square.

A hedge factor was included to accommodate nonlinear resolution over the total field of an average lens system. A 20 per cent increase in character field size was allowed resulting in a total family size of .200" x .200", at the focal plane of the first objective, in effect reducing the resolution requirements by 20 per cent.

The actual memory plate size was determined by the objective lens focal length and the total prism scan angle using the basic family field size as the prism minimum increment requirement. It was necessary therefore to determine the relative values of prism power before focal length of objective and actual prism power was determined.

The Optical System consists of an objective arranged as a collimator with respect to the photographic memory plate. In the self-parallel light output of the objective are a series of wedges

arranged with their active axes alternately orthogonal to each other. These axis are parallel to the respective x-y coordinates of the photographic memory plate. Following the wedge system, a second objective is used as a decollimator. A mask, placed at the focal plane of the decollimator, acts as a field stop. The mask size is determined by the basic family size multiplied by the ratio of the focal lengths of two objectives, i. e. : desired field size  $\times \frac{\text{Decol fl}}{\text{col fl}}$ . The field stop so determined is increased in size by a figure determined by the total accumulated error tolerance of the wedge system multiplied by a factor of two.

It is obvious from the system configuration that the total thickness of glass in the space between the collimator and decollimator may not be altered without detrimental effects on the resolution. This in effect established a restraint on the prism scanning system; a wedge may not be removed without replacing it with an equivalent thickness of glass. In consideration of this fact, the wedge binary system is designed to consist of replacement of a wedge of one power for one of like power oriented to result in precisely the opposite deviation of the wedge replaced. In general, in a system of this nature, it is necessary to keep the total prism power as low as possible to hold the angular field requirement for the collimating objective to as low a figure as practical.

It is necessary to review the design considerations involved in selecting objective focal length and wedge prism power. It will be recalled that a field size for each family of .200 inches x .200 inches had been decided upon. The total number of families in the basic vocabulary would approximate  $36 \times 30 = 1080$ ; requiring the same number of X-Y coordinates. Accordingly, the basic vocabulary could be contained in two plates with 540 families on each.

Since the field scan is by nature a square (the binary requirements for one coordinate system will be essentially the same as that for the other coordinate), the resulting unit requirement on each axis will be 24, equal to a plate coordinate population of  $24^2 = 576$ . A straight binary series will result in 16 unit power as the highest order wedge. This will result in 32 units available per coordinate if zero is a valid power. The straight binary arrangement is not applicable to the system because it infers that the objective would be used entirely in an asymmetrical system, i. e., the wedge scan when set for a total of zero power would, of necessity, be imaging at the field stop (one corner of the field presented by the objective) and when wedges are set for maximum power, the opposite corner would be imaged. This effect of additional power from a zero condition would result in a nonlinear, nonsymmetrical, field coupled with a large variation in astigmatism over the required field of the photographic memory plate.

It is desirable to operate the objective lens on axis with the memory plate centered, providing a symmetrical displacement about the optical center line. This tends to operate the lens within its most advantageous field angles. As mentioned previously, the maximum wedge power required should be reduced to a minimum consistent with the desired results. Therefore, the highest power wedge must be reduced to a value determined by the basic binary step, and order of operation, keeping in mind the fact that no zero power wedge is permitted. The central fields will therefore be displaced from the central plate coordinates which intersect the lens optical axis, so that the sides of these fields are tangent to the basic plate coordinates. Considering a family field side dimension as a unit pitch, the order of placement of the center line coordinates of a

family field with respect to the plate centerline coordinates would be  $1/2, 1\ 1/2, 2\ 1/2, 3\ 1/2, \dots, N\ 1/2$ . Accordingly, the equivalent prism power in either of the coordinate planes must provide a step change in a like manner providing a unit pitch equal to the difference between these positions. If we assume the first increment from the center to be one unit then the next is + 2 units, etc.; all even increments in the binary count disappear including zero. This count must be accomplished in both the positive and negative directions with respect to the concerned coordinate. Thus the total number of steps on a coordinate in the system will always be an even number. The binary order selected will be 1, 2, 4, 8, 10. The last value 10 departs from the normal binary system because of the desire to keep the total prism power to a minimum.

The values available are + 1 or - 1, + 2 or - 2, + 4 or - 4, + 8 or - 8, + 10 or - 10, resulting in the scalar values shown below:

TABLE 3.1 1, 2, 4, 8, 10 BINARY

<u>UNIT</u>	<u>BINARY VALUES</u>					<u>EQUIVALENT COORDINATE POSITION</u>
	<u>1</u>	<u>2</u>	<u>4</u>	<u>8</u>	<u>10</u>	
+ 1	+	+	-	-	+	+ 1
+ 3	-	-	+	-	+	+ 2
+ 5	+	-	+	-	+	+ 3
+ 7	-	+	+	-	+	+ 4
+ 9	+	+	+	-	+	+ 5
+ 11	-	-	-	+	+	+ 6
+ 13	+	-	-	+	+	+ 7
+ 15	-	+	-	+	+	+ 8
+ 17	+	+	-	+	+	+ 9
+ 19	-	-	+	+	+	+ 10
+ 21	+	-	+	+	+	+ 11
+ 23	-	+	+	+	+	+ 12
+ 25	+	+	+	+	+	+ 13

The chart shown is illustrative of the requirements of the binary values necessary to provide excursion of the field stop from the optical centerline of the system to the maximum deviational requirements. Thus providing a basic half field code for a single coordinate. The excursion required for the opposite direction is accomplished by reversing all signs within the chart. Using a binary order of 1, 2, 4, 8, 8 in Table 3.2 in both plus and minus values results in a maximum total capacity of 9216 characters 24 x 24 x 16 per plate.

TABLE 3.2 1, 2, 4, 8, 8 BINARY

<u>UNIT</u>	<u>1</u>	<u>2</u>	<u>4</u>	<u>8</u>	<u>8</u>	<u>EQUIVALENT COORDINATE POSITION</u>
1	-	-	+	-	+	+ 1
3	+	-	+	-	+	+ 2
5	-	+	+	-	+	+ 3
7	+	+	+	-	+	+ 4
9	-	-	-	+	+	+ 5
11	+	-	-	+	+	+ 6
13	-	+	-	+	+	+ 17
15	+	+	-	+	+	+ 18
17	-	-	+	+	+	+ 19
19	+	-	+	+	+	+ 10
21	-	+	+	+	+	+ 11
23	+	+	+	+	+	+ 12

In order to show capability of a 30,000 word capacity, the number of plates would have to be four giving a maximum capacity of 36,864 characters. The 1, 2, 4, 8, 10 binary (Table 3.1) results in 32,448 capacity in three plates (26 x 26). This binary system was adopted because fewer plates were required.

It was known that the effect of a wedge system such as that being designed here would result in a nonlinear effect across the field. Therefore, an area assignment for each family image was assumed to be approximately .230 inch square. A spacing factor of .020 inch was allowed between families resulting in a nominal unit field spacing of one quarter of an inch. Assuming a 12 inch focal length lens resulted in an  $15^{\circ}40'$  half angle which appeared to be a reasonable field angle for a good objective.

The prism power resulting from this lens selection and linear field was 2.5 prism diopters per unit pitch on a coordinate. The diopter values assigned to various stations become therefore:

1.25 Diopter	=	(1/2 unit pitch)	=	Binary 1
2.5 Diopter	=	(1 unit pitch)	=	Binary 2
5.0 Diopter	=	(2 unit pitch)	=	Binary 4
10.0 Diopter	=	(4 unit pitch)	=	Binary 8
12.5 Diopter	=	(5 unit pitch)	=	Binary 10

Prisms such as the wedge used in this system will refract or break up the light passing through them, into its basic elements or component frequencies resulting in a color spectrum. It is therefore necessary to achromatize the wedges to avoid a multi-colored display. Lens power is to be entirely absent and specific derivation is to be maintained within fairly accurate values. For example, if we assume that all tolerances can accumulate in one direction for a particular binary position and the opposite condition can occur on the very next binary selection, a total error of 10 x tolerance will be apparent on field stop centerline registration.

The system we are discussing actually consists of two binary systems interleaved in order to achieve orthogonal coordinate scan positions. The total maximum tolerance to be

expected from coordinate position to coordinate position on any diagonal will be 1.414 multiplied by the basic unit tolerance applied to the wedge power.

At this point, the parameters of the optical elements of the machine had been decided on and mechanical design was begun.

The individual wedge elements were designed and let for manufacture.

The lens-wedge system scheme was programmed and the combination was run through the IBM 650 Computer in order to determine actual physical locations for the centers of the individual family groups. This was necessary for preparation of the memory plate masters.

The results of this computation indicated that the contribution of the wedge system to distortion in the field of a highly corrected linear field lens resulted in an excessively large photographic plate memory due to extreme pin cushion effect.

The excessive coordinate positional displacement from the theoretical inferred that extreme image distortion could be expected from the system resulting in a probable severe loss of resolution such that the memory plate would require back-photographing.

At this point, mechanical design had progressed to the level of determining memory plate handling methods (means of changing plates) in order to show 30,000 ideograph capability. It was therefore necessary to design and fabricate a special lens to operate with the wedge system and mechanical design. In order to accommodate the mechanical design, the lens required a back focus ranging between 10 1/2 inches to 13 inches (preferably the longest practical) for clearance of the plate handling mechanism. The design specifications resulted in a reverse telephoto design (10 inch Focal

length  $f/5$  with a 13 inch back focus). This lens design coupled with the wedge system resulted in a reasonable size memory plate and saved mechanical redesign effort.

The decollimator, field stop, and projector lens group were designed as one assembly. This consisted of a 3 inch,  $f/1.9$  Wollensak Raptar Lens used as a decollimator, a field stop disk with a .070 square hole in its center and a 6.5 mm.,  $f/2.8$  Wollensak Cine Raptar lens used as the final projector.

### 3.4 Mechanical Requirements

The mechanical requirements of the various portions of the system are dictated by the optical system exclusively. Obviously as in any optical system, alignment, and preservation of alignment of the various elements were of prime concern.

Although the system is a breadboard in the sense that elementary design without consideration for the film refinements has been applied, it was necessary to rigidize the basic structures sufficiently to insure preservation of optical alignment once it has been set.

The framework was made of  $3/8$  inch plate aluminum, such that it formed a box large enough to support the optical members and their driving and mounting mechanisms.

This frame as such provides the mounting surfaces for the plate changing mechanism and its drive, the prism select and drive mechanisms, the motor and belt transmission, the collimator-decollimator projector and illuminating system. This arrangement permits mounting the whole system in any desired attitude.

### 3.4.1 Wedge Select System

The first item in the mechanical system to be designed was the wedge (prism) drive and selecting mechanism.

Several basic mechanical systems were considered for positioning the wedges in the system. One method consisted of mounting a pair of wedges and a revolving arm which would sweep through a small angle (sufficient to allow alternating two wedges in the same physical position). The second method mounted the wedges on a member which slides along fixed rails accomplishing true translation.

After due consideration, the first method was discarded. This is due to the fact that the optical properties of the wedges result in activity in one plane only. The resultant deviation is insensitive to most errors of position except rotation about the optical axis. It is immediately apparent, all things being equal, that the inertia and therefore the power required of the Prism translation system would be less than that of the rotating system. The rotating system also is inherently difficult to align optically and, most important, it is difficult to maintain alignment over a long period of operation.

On the other hand, the translational system inherently is capable of designing to a lower total mass in the moving system and is optically immune to minor changes in the mechanical arrangement of drive mechanism and stops. The translational system has a natural mechanical reference to which all optical elements may be aligned. Therefore basic adjustments are more likely to be preserved once they are made.

The Prism mounts were designed with a flange and notch so that they could be mounted on the slides and rotated by using a

special tool which engage the notch while pressing against a banking point machined into the slide bearing yoke.

The slides themselves were made of sheet metal with two yokes containing two oilite sleeve bearings at each end. The bearings within the yokes were then line-reamed resulting in a four point bearing suspension system. The rails were simply 1/8 inch diameter precision drill rod cut to the proper length to fit the slide housing. Tolerances within the slide and housing rail system were held fairly tight so that vibration and shake of the individual slides were prevented. The system of four point support is not the best approach to the problem from the manufacturing standpoint, but it results in an extremely rigid mount with very light members, obviating any possibility of rotation around the optical axis.

The time allotted for changing the binary wedge arrangement was to be approximately 1/5 of a second or .200 milliseconds. After considerable deliberation and design study, a simple electro-mechanical system was decided upon.

Drives such as direct-coupled solenoids, air cylinders, hydraulics, all required excessively complicated subsystems or had a completely useless force-stroke curve.

A swash plate crank pin mechanism was used as the basic drive mechanism. A drive bar was arranged to selectively push the slides to the desired positions. This bar carried all the slide driving escapement mechanism -- the drive bar was coupled to the swash plate crank pin through a connecting rod. The coupling was accomplished at each end of the drive bar, thus requiring two cranks. The cranks were coupled to each other in exact phase with a timing belt which also maintained timing to a one revolution input clutch.

The resulting time-displacement curve for the drive system is essentially harmonic which is close to ideal for the speed that the mechanism is expected to operate at. The design input speed at the one revolution shaft was between 360 - 400 rpm giving a total cycle time of 150 to 170 milliseconds or slightly under 1/5 second.

Ten escapement solenoids were built into the frame of the mechanism which carried the crank drive bar assembly. Each of these solenoids operates the escapement for one slide. A separate solenoid energized the one revolution clutch.

A mechanical, two-ball interposer level resets the mechanical parts of the drive bar escapement to the normal position at the end of the cycle. The control system has by this time cut off all power to the escapement system completing reset of the system to the normal condition.

The photographic memory plates were mounted in a six-sided drum. Three alternate sides accommodate the photographic plates. The remaining three sides provide a viewing window for the objective lens. Thus the lens "looks" through the interior of the drum to the opposite side. This arrangement permits a fairly compact design for the over-all mechanism. The end of the "drum" is provided with three detent dogs adjacent to the window. These dogs are engaged selectively by a detenting hook, in accordance with the control system command.

A rubber tire "Puck Drive" is used to rotate the drum from one position to another. The drive is engaged and disengaged by the action of a cam surface on the detenting hook. This arrangement prevents accidental engagement of the drive unless the detent hook is disengaged. It also insures that the drive is disengaged should hook power fail, or should the logic system for plate selection fail.

The only optical element missing from the system at this point was the illumination system. It was determined that a system with an aperture of 10 inches or better was required to illuminate the 7 1/2 inch x 7 1/2 inch plate. Since this system was not considered critical, a Fresnel lens of 14 inch focal length and 14 inch diameter was used in a doublet resulting in a 7 inch focal length lens. This lens constructed of plastic was easily cut to a size commensurate with the plate requirements.

The application of the above described Fresnel lens allowed for an extremely short "condenser" system. Permitting the lamp to be close, physically, to the memory plates. In turn, this shortened the whole system and allowed packaging in a relatively small container.

#### 3.4.2 The Control System

It was necessary that the control system provide a minimum of operator complications. Therefore, the panel controls were limited to the normal Flexowriter keyboard plus a simple On-Off power switch, and two additional switches -- one labeled "Error Cancel," and the other "Page Two."

In consideration of the method of operation of the system, it is obvious that these two controls are necessary. It will be remembered that operation of the system depends on comparison of characters presented on the view screen, to originals in printed form. This comparison is only possible after keyboard entries are made into the system. If the character required is not available in the group presented on the screen, one of two possibilities exists: either the keyboard symbols were wrongly selected or the character exists as part of page two of the family. The latter

condition could occur only if the group presented consists of 16 characters and its visual presentation indicates a page two exists, in which case the "Page Two" select button is depressed.

The former condition requires that the system be reset to the no entry condition in preparation for a new entry from the keyboard. In the case of an error selection, therefore, the error cancel switch must be depressed -- this automatically cancels all entries in the system without entering information on the paper tape.

Design consideration involved in the selection of components and operational methods within the system itself hinged almost entirely on the Flexowriter circuitry and capabilities. The requirements and therefore the system naturally divide into two basically large subsystems; the punch control system and the view control system. The division arises at the code bars within the Flexowriter due to total incompatibility of the 6 bit Flexowriter code with respect to the viewer system requirements. Synchronization of the two systems is the only inflexible requirement. (Note the block diagram Figure 3. 3).

The control circuits are arranged so that if the Flexowriter is in the lower case, its operation is normal in all respects with the exception that the upper case key will cause a shift in function.

Striking the upper case key causes a Twelve pole relay (KEC) to energize, disabling the punch and connecting the normal punch signal lines to the input transfer relays  $KT_1$  and  $KT_2$ . The input transfer relays in turn connect the punch signal lines to the Ku bank of relays in the buffer system.

The first code applied to the code lines energizes relays in the Ku bank which will remain energized due to self holding contacts. Transfer of  $KT_1$  occurs when a signal has been stored in the Ku

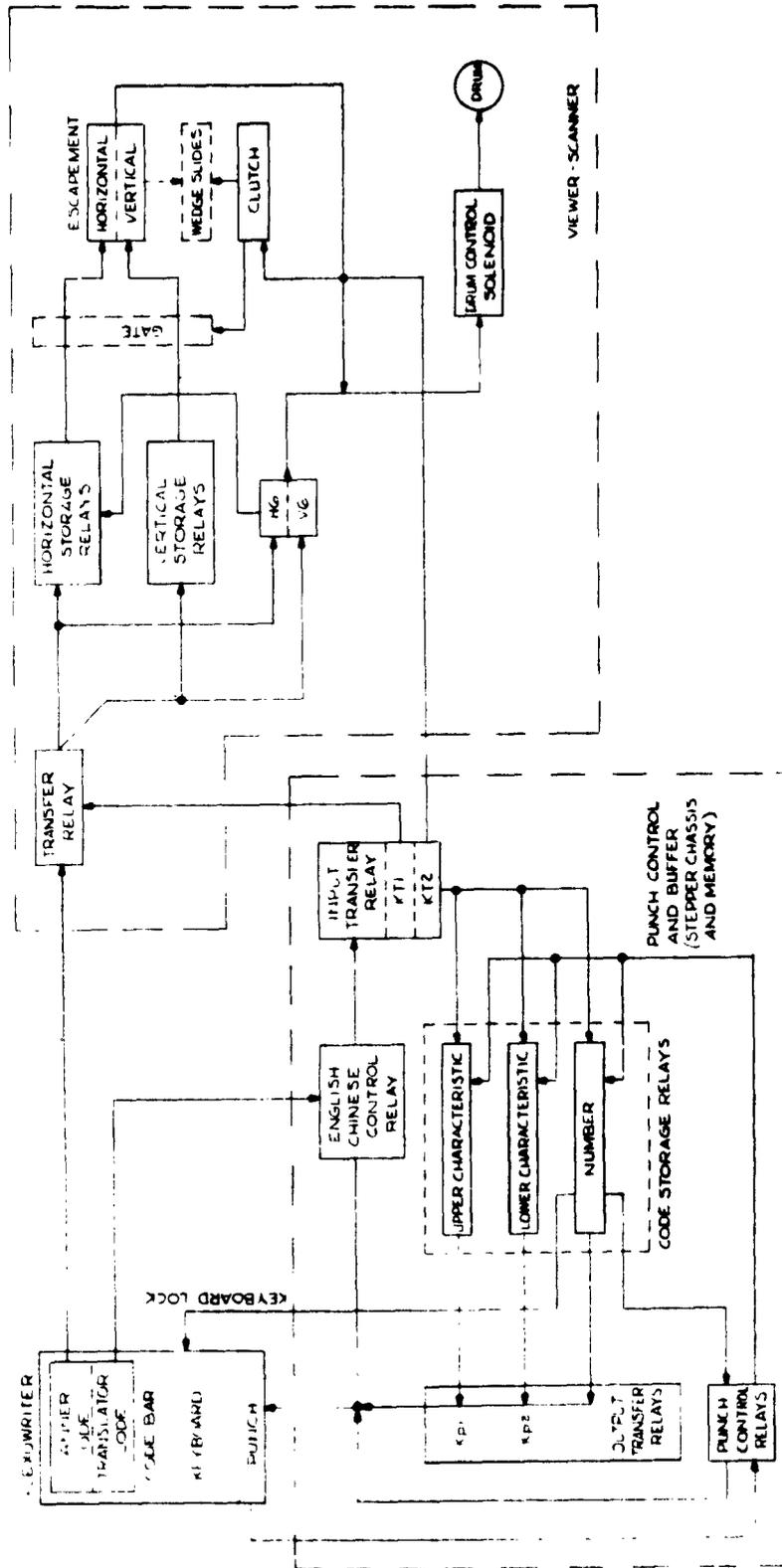


Figure 3.3 Diagram of Chinese Input Device

bank and no signal is present on the input lines. The KL bank of relays is in turn connected to the keyboard, and functions in a manner similar to the Ku bank. Transfer of  $KT_2$  occurs under the conditions outlined for  $KT_1$ , but with respect to KL bank.

The third keystroke stores a code within the  $K_N$  relays any one of which when energized causes the keyboard lock magnet to de-energize locking the Flexowriter keyboard. A ground is also completed to the read control circuit energizing KRT-1 and 2. Energization of KRT-2 initiates the generation of a control pulse to the signal control relays  $KS_1$ ,  $KS_2$ ,  $KS_3$ , only one of which may be energized at one time. Circuit logic is such that the order of energization is as listed.

Operation of a signal control relay ( $KS_1$ ,  $2$ , or  $3$ ) causes the signal output lines from the buffer relay banks to be connected through  $KP_1$ , and  $KP_2$ , to the punch magnets. Relay  $KS_1$  controls output from the KU bank,

$KS_2$  controls the KL bank and

$KS_3$  controls the KN bank

Relay KRT-2 may be energized only at the time the punch is at rest or close to the completion of its operation. A third relay KRT-3 is connected in a complimentary manner to that of KRT-2 so that while the punch is in operation the control pulse to the signal control relay is terminated by KRT-3.

After the cycle,  $KS_1$ ,  $KS_2$ ,  $KS_3$ , is completed, it is necessary to provide a punched code indicating end of word, this is injected through a fourth relay (KRS) connected to the same signal buss as the signal control relays. This (KRS) relay serves two purposes:

1. It injects the required end-of-word code signal.

2. It removes power from all relays except KEC  
(English/Chinese),

thus preparing them for the next series of code entries.

All the relay systems described above are contained on two chassis designated the Stepper and the Memory.

#### 3.4.3 Viewer Control System

As described previously, the slide drive bar mechanism is equipped with solenoids which directly control the positions of the slides. The solenoids are biased prior to the mechanical operation of the drive bar. Since the slides contain the wedges, they therefore determine the coordinates or location of the field stop.

Due to the fact that redundant coding conditions with respect to the scanner requirements arise during the keyboarding operation, it is necessary to provide the scanner wedge system and drum select mechanism with a buffer somewhat similar to the memory section of the punch control circuit described above.

The scanner buffer serves as a plate selector and signal phase discriminator, in addition to storing the horizontal and vertical scanner codes supplied from the keyboard. A six level code is applied to the scanner buffer, the first five levels are analogous to the required wedge positions and are directly connected to the respective escapement solenoids. The sixth level is connected in a grounding circuit through three micro switches, to the index control solenoid select switches located on a cam on the drum shaft. The drum rotates until it opens the grounded circuit, de-energizing the indexing solenoid and stopping at the desired memory plate.

A single six pole relay serves as the viewer transfer relay ( $KST_1$ ). This switches the keyboard scanner code output from the horizontal scanner code relays to the vertical scanner code relays. The scanner transfer relay ( $KST_1$ ) coil is directly connected to the coil of  $KT_1$  in the stepper chassis. This insures synchronization of the input signals. The horizontal scanner relays are energized when the upper characteristic is coded. The vertical scanner relays when the lower ideograph characteristic is coded.

At the operation of the input codes, the proper output to the scanner lines is energized by applying a common ground in synchronization with the punch system. This synchronization is achieved by using the ground that appears at the  $KT_2$  coil in the stepper chassis; thus interlocking is completely achieved by using the ground from  $KT_1$  and  $KT_2$ .

Completion of design of the control system permitted consideration of the type of chassis on which all units were to be mounted.

#### 3.4.4 Packaging

Because of the ability to shorten the condenser system by use of a Fresnel lens, it was found that the system as a whole could be mounted within a standard off-the-shelf cabinet. The cabinet just accommodated the system, therefore flexibility of mechanical design was limited.

The structure of the cabinet was modified to provide a set of beams forming a nest for the optical unit. The unit is readily unfastened by removing only four screws.

The weight of the typewriter when mounted on the work surface of the unit was sufficient to tip the whole cabinet over when

the optical unit was removed. It became necessary therefore to extend the base and reinforce the typewriter work surface. This was accomplished by using thin wall tubing for the reinforcing, and angle iron and aluminum plate to extend the cabinet base under the Flexowriter.

### 3.5 Coding Arrangement for Word Acquisition

#### 3.5.1 Photographic Memory Plates

The total capacity of the three photographic memory plates provided is 32,448 ideographs if all the space is utilized.

It was necessary to investigate the natural control parameters governing the family distribution versus the scanner capability, in order to determine memory plate space allocations.

By nature, the basic scanner system is a two parameter area search device, its requirements being input values for selection of the x and y coordinates of a focal plane. It is apparent that the key-boarding classification is analogous to the X, Y coordinates since the family designation is obtained by the upper geometric form and the lower geometric form. The analogy leads to the immediate conclusion that the upper characteristic or geometric form should equal the horizontal coordinate and the lower characteristic the vertical coordinate in the focal plane.

A system of space allocation is necessary to determine the code requirements for the various key applications. The following describes the methods employed:

It will be recalled that there are 36 keys on the keyboard assigned the upper characteristics and 30 assigned the lower characteristics. These will therefore require a coordinate system of 36 by 30 units.

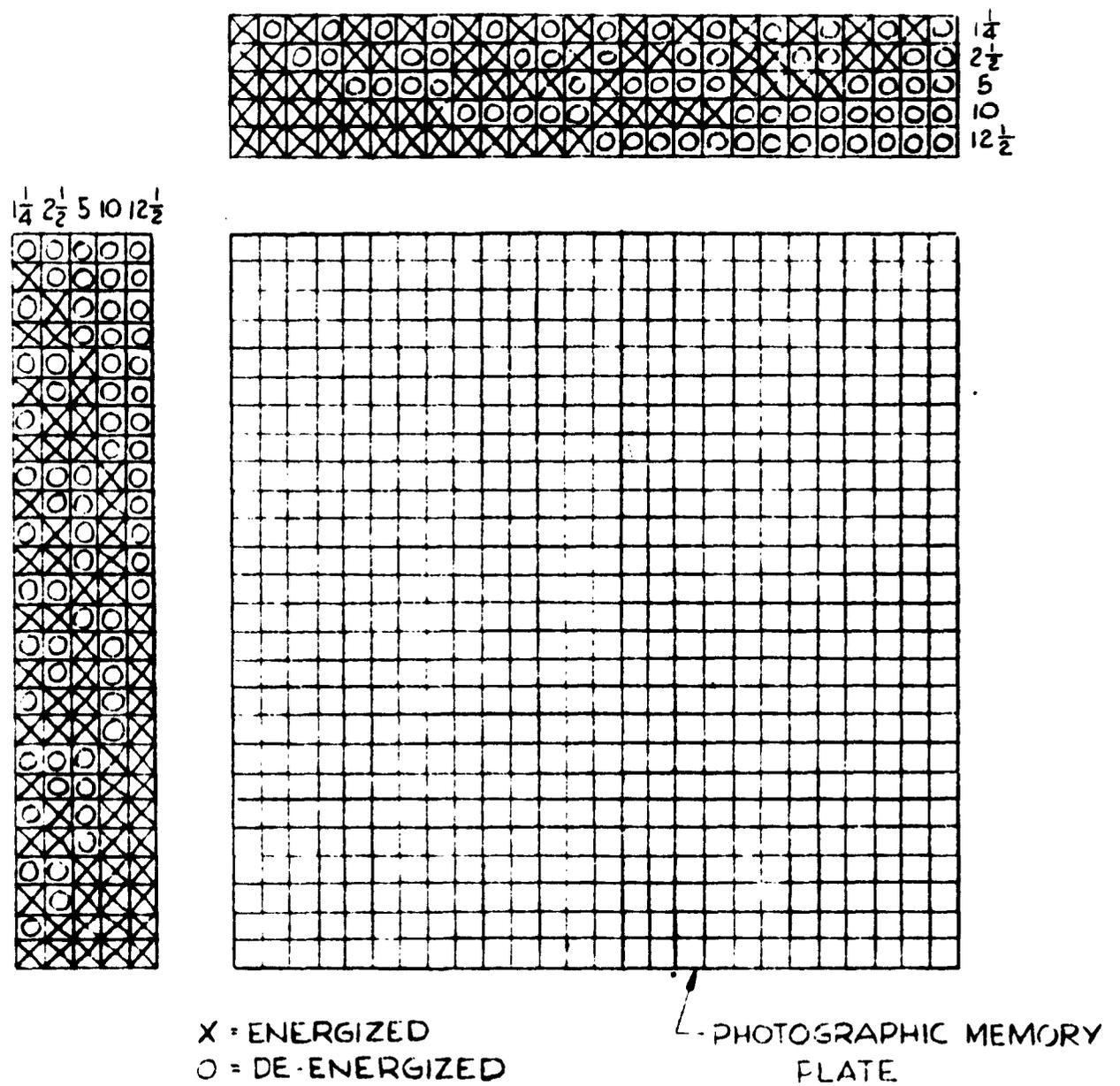


Figure 3.4 Address Codes Photographic Memory Plates

With a five-bit input duplexed in the X then the Y coordinates the scanner is capable of selecting a unit from the 26 x 26 (memory plate capacity) coordinate system. The basic bit allocation is as shown in Figure 3.4, in accordance with a symmetrical disposition of the binary arrangement shown in Table I.

If the three plates shown in Figure 3.5 are removed from the drum and placed edge to edge so that the horizontal coordinate is shown as a continuous line, it will be seen that the vocabulary is easily distributed horizontally on Plate I, units 1 to 26; and Plate 2 units 1 to 10. Unit 1 on Plate II is equal to vocabulary unit 27 and unit 10 on Plate II equals vocabulary unit 36 (see Figure 3.6).

The real problem of space allocation occurs due to the cross-hatched portion of the vocabulary area -- plate overlay shown in Figure 3.7. Vertical units 27 through 30 require the same horizontal codes as those in Units 1 to 26, and if the vertical units 27 to 30 are assigned an area on Plate II using the codes directly, there would be area interference. It will be noted from this figure that Plate III is entirely empty and therefore it can be used with a direct application of the same horizontal code applied as on Plate I, just as if Plate III were a continuation of Plate I, in the vertical sense (Figure 3.8).

In approaching the problem of extending an area such as shown above, the keyboard codes assigned remain unchanged for application to Plates I or II on the horizontal scan. Discrimination as to which plate is to be active is assigned to the sixth bit available on the keyboard code bars. Unit codes on Key 27, 28, 29, 30 are identical to unit codes 23, 24, 25, 26, with the exception that the sixth level bit is also present.

A similar discrimination in horizontal codes is used to determine Plate II selection. The scanner codes for all three plates

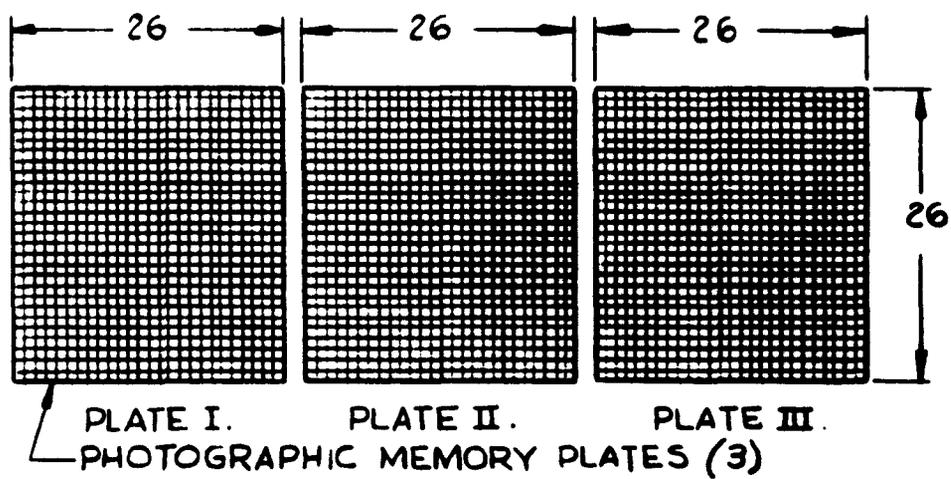
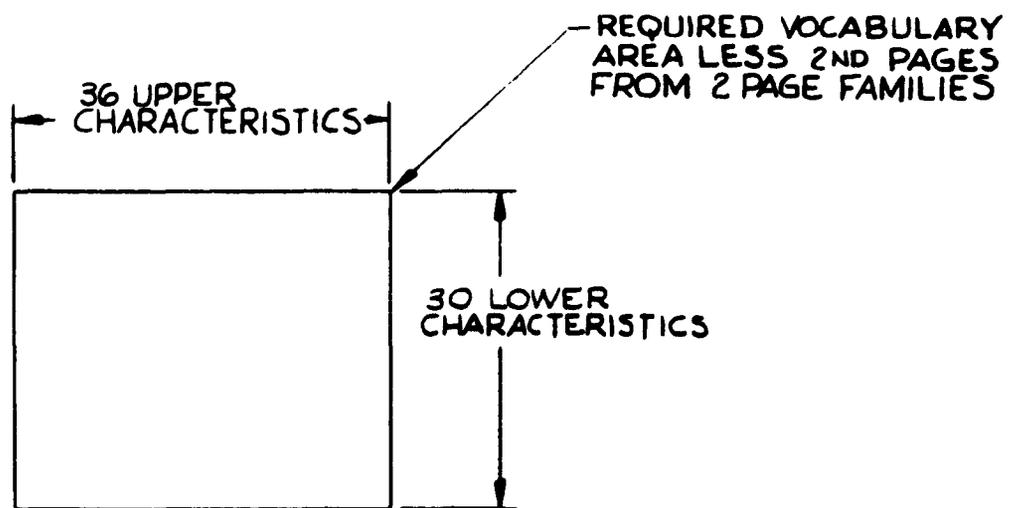


Figure 3.5 Relative Logical Space Available

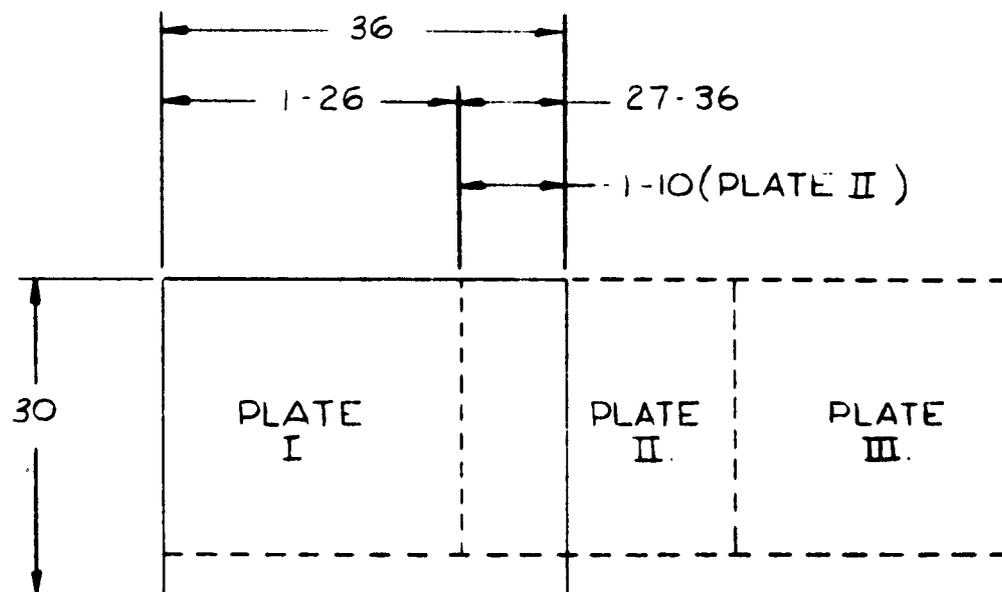


Figure 3.6 Overlay of Areas Depicted in Figure 3.5

horizontally are identical. However, the keys carrying the unit characteristic 27 - 36 are equipped with a sixth level bit.

In this manner, a simple discriminator circuit determines which of the three plates is to be scanned as indicated:

1. No sixth level bit in the horizontal and vertical scanner codes instructs the unit to scan Plate I.
2. No sixth level bit in the vertical but a sixth level bit in the horizontal means scan of Plate II is desired.
3. No sixth level bit in the horizontal but a sixth level bit in the vertical means scan Plate III.

It will be noted that the additional horizontal coordinates 27 - 36 for the vertical families 27 - 30 will occupy the same space as those continuing Plate No. 1 lines 23, 24, 25, 26. This situation is intolerable but is easily handled by application of an extension of the logic listed above.

It will be remembered that the codes determining the wedge positions (5 levels are required) is topographically symmetrical about the optical axis of the system. This means that a code calling for the right hand Unit No. 26 is a mirror image of the code calling for the left hand Unit No. 1, represented by XXXXX = 26 (Table 3.1) then 00000 = 1.

Since the extension of the horizontal classification is less than half of the Plate II units positions in the same direction, it is a simple matter to assign the scan code an inversion in the horizontal plane when a sixth level bit is present in both the Horizontal and Vertical scan codes.

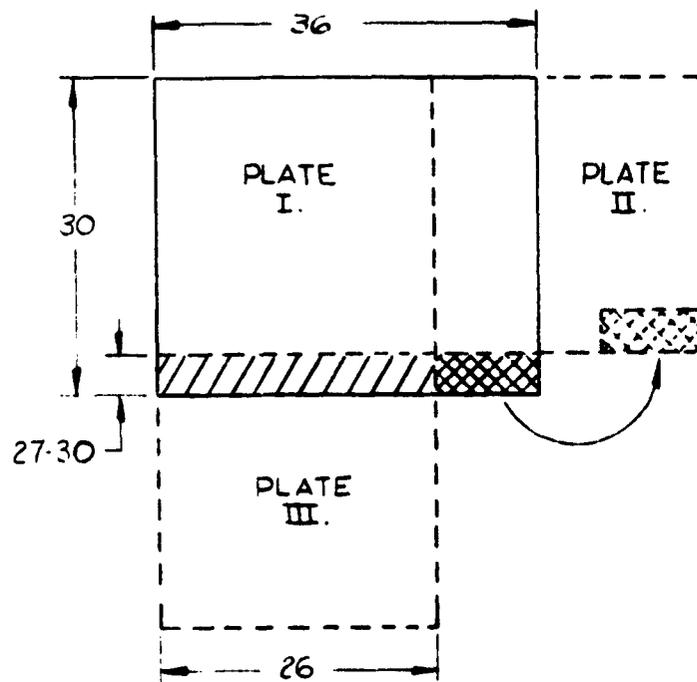


Figure 3.7 Diagrammatic Representation of Coding Problem Area

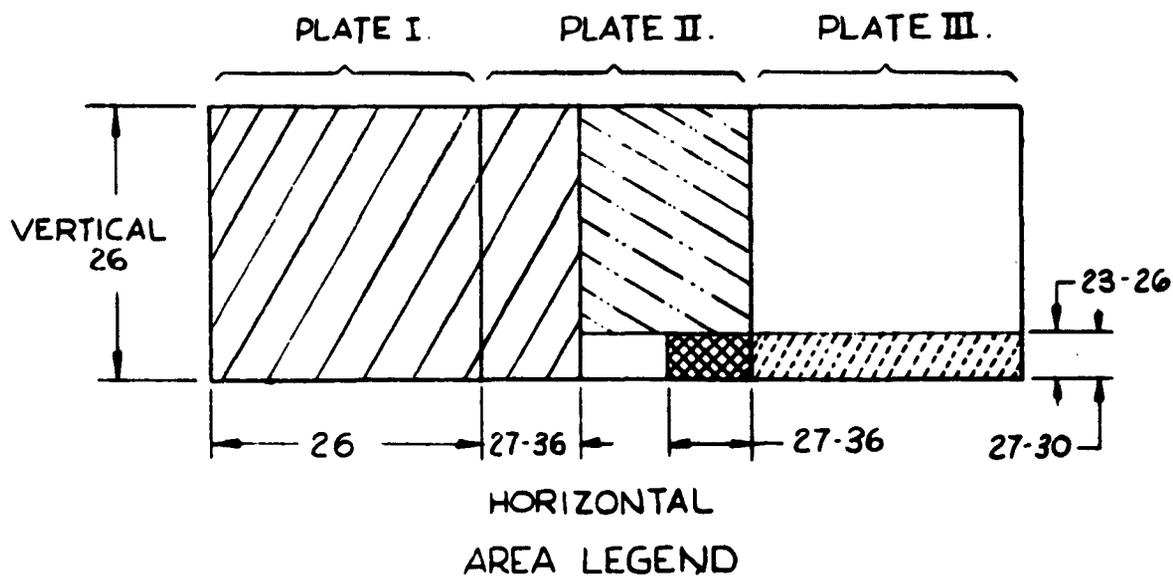


Figure 3.8 Diagrammatic Representation of Coding Problem Area

Thus Plate II has two horizontal designation schemes. When referring to the vertical code group and no sixth level bit is present, the codes for horizontal scan are 1 - 26, Left to Right; if a sixth level bit is present in the vertical code the order is reversed and 1 - 26 is read Right to Left.

Since the optical system code is invariable 1 - 26 left to right, 1 - 26 top to bottom, it is necessary to manipulate the code as fed to the scanner escapement. This manipulation, since it is simply a mirror image requirement, is easily established by reversing the phase of the output codes.

Thus the reversal of the horizontal code requirement is accomplished when both Horizontal and Vertical scanner buffers contain a sixth bit, simply by reversing the ground connection on the buffer output relay switch points. Normally, if the horizontal buffer calls for ENERGIZED relay equal to a + position for a wedge slide, the ground would be applied to the NO contact of the output leaf on the relay. A phase reversal is accomplished by changing the ground to the NC side of the contact group. (It will be noted that binary values are arbitrarily assigned for unit position rather than taking the actual decimal equivalent numbers which in this case have no real value).

Solution of the coding and plate select problem was completed at this point with the exception that families with more than one page (over 16 characters) would constitute a complication, in that a code changer might be necessary for the viewer escapement if they were allowed to fall into areas whose complements were occupied by other families. It was felt that the simplest approach to the page two problem was to allow the scanner to remain at its first coordinate position and rotate the drum to alternate positions: Plate II or III.

Since very few families (20) consisted of groups requiring two pages, it was a simple matter to assign to the first page groups that area of Plate I defined as being between 1 - 22 vertical codes and 11 - 26 horizontal codes. This automatically provides a non-ambiguous space on Plate II for the second page of these groups.

The drum could be indexed by energizing the sixth level in the horizontal code relay without changing the punch code group memory. By using this approach, the logic of the system was kept as simple as possible.

### 3.6 Problems

#### 3.6.1 Plate Preparation

The original masters used by Lin Yutang were photographically enlarged to a size necessary to provide each character with a 1 inch by 1 inch field. The family groups were re-organized so that ambiguity was removed by providing redundancy of position where there was doubt as to what characteristics an individual should be classified under. Therefore in the Chinese Input Device System a character could have two addresses.

A series of master sheets providing a 4 inch by 4 inch grid were printed up and the individual members were pasted up. An attempt was made to assemble the characters on the grids in an order approximating frequency of usage. This was done by familiarity with the language and was not based on any statistical studies. From these a master plate was made.

It was necessary in assembling the master plate to preserve family locations accurately. The family centerline positions required precise layout to within a few thousandths of an inch of

the theoretical in order to allow minimum tolerances on the placement of the individual families during the cementing operation.

Difficulty was experienced in obtaining useful prints of the masters due to the requirement of a minimum of 20 lines/mm throughout the plates, and the necessity to compromise an optimum condition for the reproduction.

A greater difficulty however, was keeping the dictionary-code relationship organized and assembling according to the pre-conceived arrangement.

### 3.6.2 Control System:

In design of the control system it was necessary to preserve operation of the Flexowriter in its normal sense, which in turn required retention of the unit circuitry intact. The system was designed as a patch-cord device, all switching and modification to be done within this cord.

As the IBM six bit Flexowriter code is not compatible with the internal coding employed by Mergenthaler, it is therefore necessary to alter the Flexowriter to supply two different codes from a single code bar, one with the standard six bit code and the other with the Mergenthaler scanner code.

The only other sources of trouble arose due to the fact that the Flexowriter power supply and signal circuits required electrical isolation from the punch buffer-scanner system in order to prevent short circuits and misoperation. Diode isolation was the most direct approach to the solution of this problem.

### 3.6.3 Optical Alignment

As noted in the mechanical description, it was necessary to

provide a reasonably accurate method of alignment of the wedge system. It was considered a practical impossibility to accurately align the wedges within the system as an assembly.

Each wedge holder was designed so that it was a slip fit in its locating hole in the wedge slide, and each holder had a reference notch used also for adjustment during fine alignment.

A prismatic ranging tool was used to obtain alignment of the wedge proper with respect to the holder-notch. In this tool, the wedge holder is positioned with its slot engaging a pin and the wedge is cemented with epoxy into the position resulting when coincidence of the reference line to the alignment telescope reference line is achieved. For the horizontal elevation wedges, the pin is in the horizontal plane; for the vertical wedges, the pin is moved  $90^{\circ}$  to the vertical plane. (All alignment is with respect to the horizontal reference line).

The basic alignment described above positions the wedge within its holder to an accuracy estimated at  $\pm 1/4^{\circ}$  which is sufficient to allow final adjustment in the slide assembly without recourse to special methods.

The slide holders, as mentioned above, have a natural reference line built in. This is a line established by the oilite bearings on which the slides mount and operate. It is a relatively simple matter to align the wedges within the slides by optical autocollimation. A fixture is arranged, on a surface plate, to hold a pair of rods similar to the ones on which the slides operate parallel to the master surface. The rods are referenced to the alignment telescope so that their plane is perpendicular to the optical axis of the scope. An accurate roof mirror is placed on the optical axis with its axis perpendicular to the optical axis. This condition is established by autocollimation.

Setup of the Telescope Roof Mirror combination must be done very accurately, as all alignment depends upon it.

A property of the roof mirror is that all rays of light entering at right angles to the roof axis are returned parallel to themselves; rays entering with a component parallel to the roof axis are reflected in this plane as if they were reflected from a plane mirror. (see Figure 3.9).

For the Horizontal deviation wedge-slide assembly it is necessary to align the wedges so that they contribute no wedge power to the vertical scan system, therefore, the roof mirror is arranged with its axis perpendicular to the master surface (rod bearing axis taken as the horizontal deviation axis). This condition is ascertained by rotating the collimating telescope through a small angle with respect to the roof mirror, keeping the axis of the "scope" parallel to the master surface. When the return image remains in coincidence with the collimator reticule, the mirror is perpendicular to the surface and the mirror is perpendicular to the optical axis of the collimator.

A plane parallel mirror is now mounted on the slide mounting rods in front of the roof mirror, and the telescope is auto-collimated to the plane of the rods without disturbing the horizontal parallelism of the telescope to the reference plane. The mirror is removed and a wedge slide assembly is placed on the rods. The wedge holder is placed on the optical axis of the collimator by sliding the assembly back and forth. The wedge holder is revolved on its axis by using a special tool until the return image coincides with the reticule. The holder is then locked in place with two mounting screws. The process is repeated for the second wedge holder in the slide. This automatically provides an optically active axis

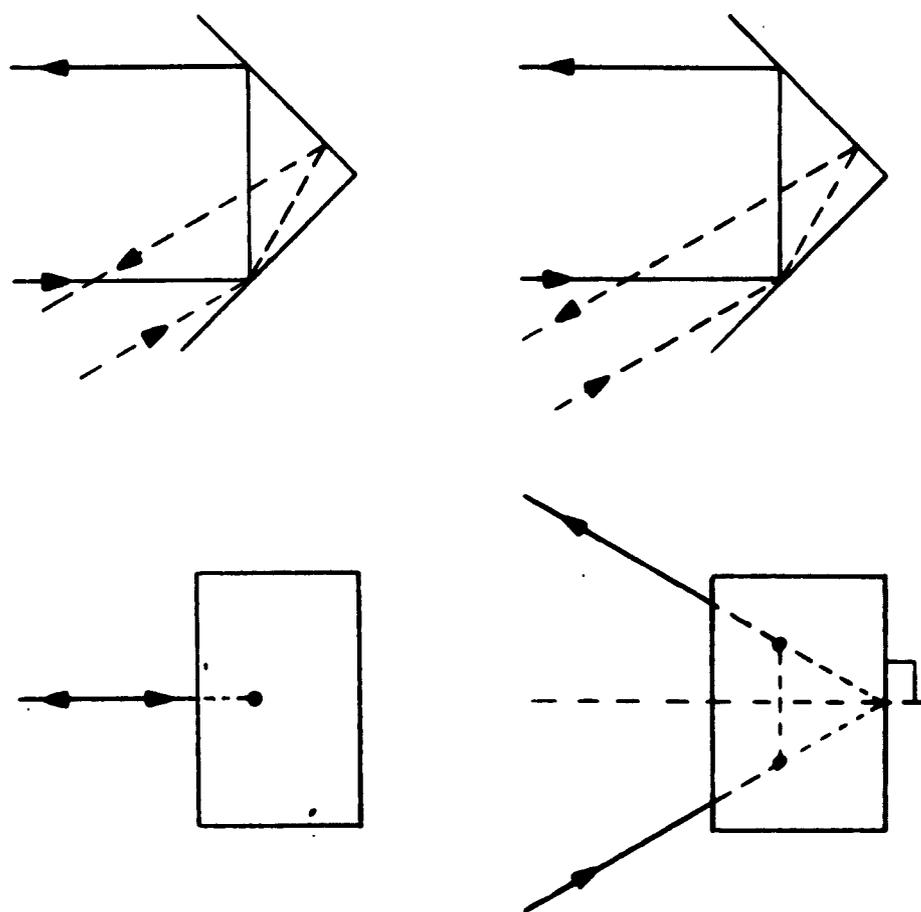


Figure 3.9 Roof Mirror

parallel to the mounting rods, with zero power in the vertical scan. The accuracy of this alignment is only limited by the resolution of the collimator mirror and wedge which is on the order of seconds of arc. The procedure is repeated for all the horizontal scan slide-wedge assemblies.

The vertical wedge alignment requires a change in the position of the roof mirror -- the axis of the mirror must now be parallel to the slide mount rods or, more exactly, at right angle to the previous orientation. The alignment scope position is not disturbed, and the roof mirror is rotated through  $90^{\circ}$  and again autocollimated. The angle of the axis of the collimator is changed slightly with respect to the master surface and coincidence of the return image is noted. NOTE: Elevation of the return image must remain the same regardless of angle of the collimator to the master surface. A deviation left or right simply indicated nonperpendicularity of the effective mirror plane to the optical axis. Zero deviation in the vertical plane indicates parallelism of the mirror axis to the master surface.

When the alignment of the roof mirror has been determined, the collimator is again autocollimated to the mount bars and all wedges for vertical scan are aligned as in the horizontal scan system.

The basic alignment outlined above establishes a precise orthogonal optical X, Y coordinate system accurately referenced to the wedge drive and selection mechanism.

#### 3.6.4 Lens Elements

The design of the mounting system for the various lens elements is such that they are intimately related to the wedge system.

Master mounting surfaces for the main objective-collimator and decollimator field stop projector lens assembly are made part of wedge assembly frame. Therefore all important optics are contained within one assembly and are self-referenced as a unit.

This assembly is now placed within the main frame of the unit and mechanically oriented in order to center the optical axis on the drum window (where photo plates will be mounted) so that the axis is perpendicular to the window plane.

Basic alignment of the system is now completed.

The photographic memory plates are now placed in their respective positions and the system is focused to collimation on these plates for the main objective. When this has been accomplished, registration of the field stop with the four corner families on a plate is accomplished by moving the plate on its mounts. After the registration is accomplished, the memory plate is locked into position.

All three photographic plates are registered in the same manner.

Because of some small play in the decollimator lens barrel, a final adjustment for best compromise of position is accomplished by adjusting the field-stop positioning clamps on the main unit body.

#### 3.6.5 Illumination

A Fresnel lens was assembled as previously described and mounted below the drum on the main assembly. Below this a mirror of appropriate size has been mounted at  $45^{\circ}$  to the optical axis. A 500 watt EDK bulb was mounted at the proper focal distance from the Fresnel lens in order to image the lens filament at the aperture stop of the wedge system.

It was found to be impossible to illuminate the whole field sufficiently to make a reasonable presentation on the screen with a standard bulb and the condenser system described. The factors involved several optical problems (enumerated below), tending to reduce the light at the screen for the extreme family groups (edges and corners).

- a. Spherical aberration of the light-source imaging system.
- b. Loss due to apparent reduction of system aperture at extreme scan positions.
- c.  $\text{Cos}^4$  law<sup>1</sup> which relates to the losses across a field as the field angle is increased (both the field angle of the condenser and the field angle of the objective are involved here).
- d. The wedge system is inherently a spectral analyzer and it is difficult to achieve absolute color correction for all binary combinations of wedges. The worst color effect occurs at the extreme scan position.
- e. The relative aperture of the system is unbalanced across the field, that is, it is an asymmetric system, where the effective central aperture (that aperture composed of a symmetrical bundle about the axis) is considerably reduced in size.
- f. The over-all system is operating at a relative aperture of approximately  $f/22$ .

It was found that these all tend to aggravate the lighting problem to an extent that made it impractical to attempt a solution

---

<sup>1</sup>  $I = BW' \text{Cos}^4$

through redesign, since it would require complete re-manufacture of the entire system. The above would only be possible if there were any solution at all for the field with which the system were concerned.

### 3.7 Quality Control

3.7.1 Originally a wedge system was let for manufacture allowing deviations of  $\pm 0.1$  Prism Diopter. It was felt that the tolerances on a set of wedges would tend to average out near the mean. After assembly, it was found that the wedges all tended to higher power and were also approximately twice the value of the specified tolerance from nominal. Needless to say the field would not register properly. These first wedges also suffered from excessive lens power and faulty color correction.

3.7.2 The lens which was designed for the system suffered from extreme astigmatism and bad color correction at the edges of the field.

3.7.3 New glass elements for the wedges were ordered with much tighter tolerances with respect to the three problems mentioned above. Close liaison with the vendor revealed that the method of cementing (specified by MLCo.) caused excessive errors in lens power and deviations.

A third group of wedges was ordered from another vendor which was just within tolerance on deviation, and did not suffer excessively from lens power problems. This group still presented some lens power on one or two wedges which is most noticeable for certain combinations. There were no wedges of better quality available from the extras ordered.

It was necessary to set up a special inspection procedure for acceptance of these wedges in order that all units placed in the system could be matched. (It was found that tolerance accumulation due to pairing of the wedges resulted in excessive "dance" of the projected image within the field stop, sometimes making it impossible to achieve complete registration of all families).

Pairing wedges for equal deviations resulted in almost perfect registration per quadrant (13 units x 13 units) causing the tolerance accumulation to be most apparent at the centerline of the photographic memory plates.

Although specifications for the elements had been prepared for a considerable period of time, certain manufacturing difficulties appeared and caused us to execute a recheck on the theoretical versus the actual performance of each wedge with reference to its nominal power.

It was found that the glass for the 12.5 P.D. wedges (theoretical value) were actually specified per manufacturing dimensions as 12.7 Diopter wedges. Since the physical deviation for the system was specified as 12.5 Diopters, these wedges were assembled to produce 12.5 Diopter. The difference between the glass (12.5 P.D.) and the system (12.7 D.P.) constituted a built-in error resulting in slight inaccuracies of color correction. The displacement calculation (X-Y) coordinate positions on the photographic plate were based on the 12.7 Diopter wedges resulting in basic mismatch of the photographic plates with the system.

3.7.4 A new lens was designed and fabricated to remove the excessive color and astigmatism.

3.7.5 The new lens design resulted in the necessity for new calculation for the coordinate positions which in turn resulted in the necessity of preparing new photographic masters for the vocabulary.

3.7.6 Because new masters were required, a three step technique was restored to, allowing use of the negative from the camera as directly applicable within the system. In this way over-all resolution was improved.

It was necessary to photograph each family group on paper directly to the 1 inch x 1 inch scale. These were in turn pasted up and then photographed onto the glass plates to the proper scale. This work also afforded the opportunity to touch up individual ideographs so that stroke weight was approximately equal, thus allowing more accurate exposure and developing techniques permitting the best resolution from the reproduction methods.

#### IV. CLASSIFICATION OF CHARACTERS

One of the main aspects of the input problem is the difficulty of classifying Chinese characters. This is especially emphasized when the aim is for the education of non-Chinese personnel to operate any input mechanism for Chinese-English mechanical translation.

Chinese scholars have long recognized the unique difficulties of clear classification in an ideographic language such as Chinese and have struggled with this problem for decades. However, no really satisfactory solution has yet been proposed. We review below, in brief, the major methods that have so far been used and discuss their merits and shortcomings.

To begin with, we can divide the over forty schemes that are in existence into two major families: 1. Classification by Shape, and 2. Classification by Sound.

We will not concern ourselves with methods of classification by sound, but will direct our attention to methods of classification by shape.

Methods of classification by shape can again be subdivided into these major groups: 1. Classification by Radicals, 2. Classification by Stroke Types, 3. Classification by Graphical Recognition, and 4. Classification by Number, which are discussed in the following paragraphs.

##### 4.1 Classification by Radicals

The primary example of this system of classification is the radical stroke system as employed by the Kang-Hsi dictionary. This is the classical method of classification and has been employed by scholars since the Ming Dynasty. A set of 214 radicals are defined and all Chinese characters are grouped as belonging to one of the 214

radicals. To use this method one needs to be familiar with all 214 radicals and to be able to determine at once to what radical the character in question belongs. After determination of the radical, the number of the strokes in the character outside the radical is counted. This further narrows down the number of characters within the group. Finally, a list of characters possessing identical number of strokes is consulted to determine the correct character. As can be seen from the above description, this method is cumbersome and it would be very difficult to train non-Chinese personnel to employ this system with any degree of flexibility and speed.

#### 4.2 Classification by Stroke

To avoid the pitfalls of the system of classification by radicals, a new approach was taken by other students of Chinese character classification.

In a sense, a Chinese character may be called alphabetic if one considers each type of stroke as an alphabet. If one follows this train of reasoning, one may claim that each Chinese character is no more than a combination of "stroke-alphabets."

Thus, we can rationalize the Chinese vocabulary into 40 basic strokes, combinations of which generate all characters. These 40 strokes may be further classified into six types -- dot, line, angle and arc, etc. Chinese characters can then be defined by the order in which these strokes appear.

However, objections to this method can immediately be raised. First, in Roman alphabetical languages, one alphabet succeeds another from left to right without exception, while this is hardly the case with Chinese characters, for the strokes follow each other in all directions. Second, there is no physical order

of writing of the strokes, although by common consensus, certain stroke orders are generally followed in writing certain characters. Yet, there are no hard and fast rules whereby the correct order can be determined. Thus, a very common character 門 when written by four different people -- each a native Chinese -- may have the stroke orders arranged in four different manners. With these objections in mind, we can apparently ignore this system for input purposes.

#### 4.3 Classification by Graphical Recognition

The basis of this system is the division of Chinese characters into geometrical areas and the identification of elements within each area by numbers.

One popular system utilizing this concept is the four-cornered classification system invented by Yung-Wu Wang. Briefly, this system may be summarized as follows:

1. The various strokes employed in writing Chinese are classified into ten types denoted by numerals from 0 to 9.
2. Any character is identified by the strokes in the four corners of the Chinese character, the order of identification being: 1. Upper Left, 2. Lower Left, 3. Upper Right, 4. Lower Right.

The strokes in each corner are identified by its numeral, thus generating a four-digit code. There are other rules which make this system more sophisticated, but the two enumerated above are sufficient to represent the general system.

This system is useful, but suffers from the duplication of code numbers by various characters. A fifth digit has been appended when such duplications exist, in an attempt to allow for unique identification of each character, but it is still not possible to provide unique coding for each and every Chinese character.

#### 4.4 Classification by Number

With the introduction of Western telegraphy into China, a previously unknown problem arose, namely, the efficient translation of Chinese characters by telegraphy. To combat this problem, the telegraphic code book was compiled.

Basically, this is no more than the utilization of the radical stroke system, arranged in the usual manner and numbered serially from 0000 to 9999. However, not all numbers are used consecutively. Later, desire for a greater vocabulary caused the addition of another section into this code book, again arranged by radical stroke order, and with each subsequent revision, new sections have been added. Thus, despite its resemblance to the radical stroke order, the fact is that a stranger to this code book would have to rely on radical stroke order and trial and error to discover the code number, whereas experienced telegraphers have memorized the equivalent code for each character. Therefore, we can safely say that this has evolved into a new system where a unique number is assigned to each character, and where there is some semblance of order, but where no definite rules can be formulated for locating a particular character.

## V. ESTIMATE OF THE TOTAL NUMBER OF CHARACTERS EMPLOYED IN CHINESE SCIENTIFIC TEXTS

The task of determining precisely the total number of characters used in modern Chinese scientific writing is a difficult one due to the many factors that have to be taken into consideration. Some of these are:

1. Stylistic difference among authors
2. Transliteration differences among authors
3. The lack of a rigid standard for terms used in scientific writing
4. The broadness of the term "scientific texts"

An educated estimate of such a number can nonetheless be made and the figure thus obtained serve as a valuable guide-line for our purposes.

During the course of this study, a number of Chinese-English, Chinese-English-Russian, Russian-Chinese, Chinese-Russian, special terminology dictionary were examined. Table 5.1 lists 10 special terminology dictionaries. Each entry in these dictionaries (either a word or a phrase) on the average four characters. Thus, for the ten dictionaries, we would have approximately 700,000 running characters. This at first glance seems a rather high figure.

However, individual characters are very often duplicated for use in the entries and form a greater part of the running characters. Thus, on a single page one character may appear as much as seventeen times (the character 石 on page 31 of 矿物学名词). With this fact in mind, we would therefore estimate that a dictionary or an input device with a vocabulary of 8500 to 12,000 characters will be amply sufficient to handle all modern Chinese scientific, journalistic and literary writings.

**Table 5.1 List of Special Term Dictionaries**

Special Term Dictionaries	Date of Publication	Fields of Discipline	Total Entries Approximately	Average Characters Per Entry	Total Estimated Characters
1.	February, 1957 Academia Sinica	Mineralogy	8,576	4	34,304
2.	September, 1957 Academia Sinica	Physics	28,720	4	114,880
3.	1958 Academia Sinica	Astronomy	4,048	4	16,192
4.	March, 1958	Meteorology	8,160	4	32,640
5.	April, 1955		10,480	4	41,920
6.	China Commercial Press, April, 1954	Railroad and Highway Engineering	8,700	4	34,800
7.	Academia Sinica	Chemical Engineering	16,000	4	64,000
8.	1959	Construction Engineering	45,000	4	180,000
9.	1959 Academia Sinica	Organic Chemistry	10,000	6	60,000
10.	1957 National Defense Industries	Aeronautical Engineering	40,000	4/5	200,000

## VI. LEXICAL PROCESSING

### 6.1 Preparation of Input Material

Two articles selected as study articles were photographed, and several copies are reproduced, one copy being sent to the Sinowriter operators for the preparation of the input tape.

As the Sinowriter has not been fully developed at the stage of input preparation, the input tape was prepared by manual simulation of the operations of the Sinowriter. A coding book was prepared which is a replica of the code forms stored in the Sinowriter. The operator codes the first two bytes of the code as on the Sinowriter and sets it down on paper. At this point, instead of looking at the screen for the third code, as would be done on the Sinowriter, the operator checks the code book, which is arranged alphanumerically by the first two bytes of the code. The third code letter is then written down to complete the coding of the character.

When the entire article has been coded, the operator types it on a Flexowriter. The resulting tape is read as the input tape.

### 6.2 Detection of Input Errors by Means of a 704 Unidirectional Single-Pass Translation

After the input tape has been prepared, it is checked for errors. This is accomplished by attempting a 704 unidirectional single-pass translation. This pass allows for detection of errors not only in the input tape, but also in the Master File. The operation is described in Flow Chart 6.1.

### 6.3 Master Reference File

#### 6.3.1 Description of MRF Card

For working purposes, and for the generation of Search Input State

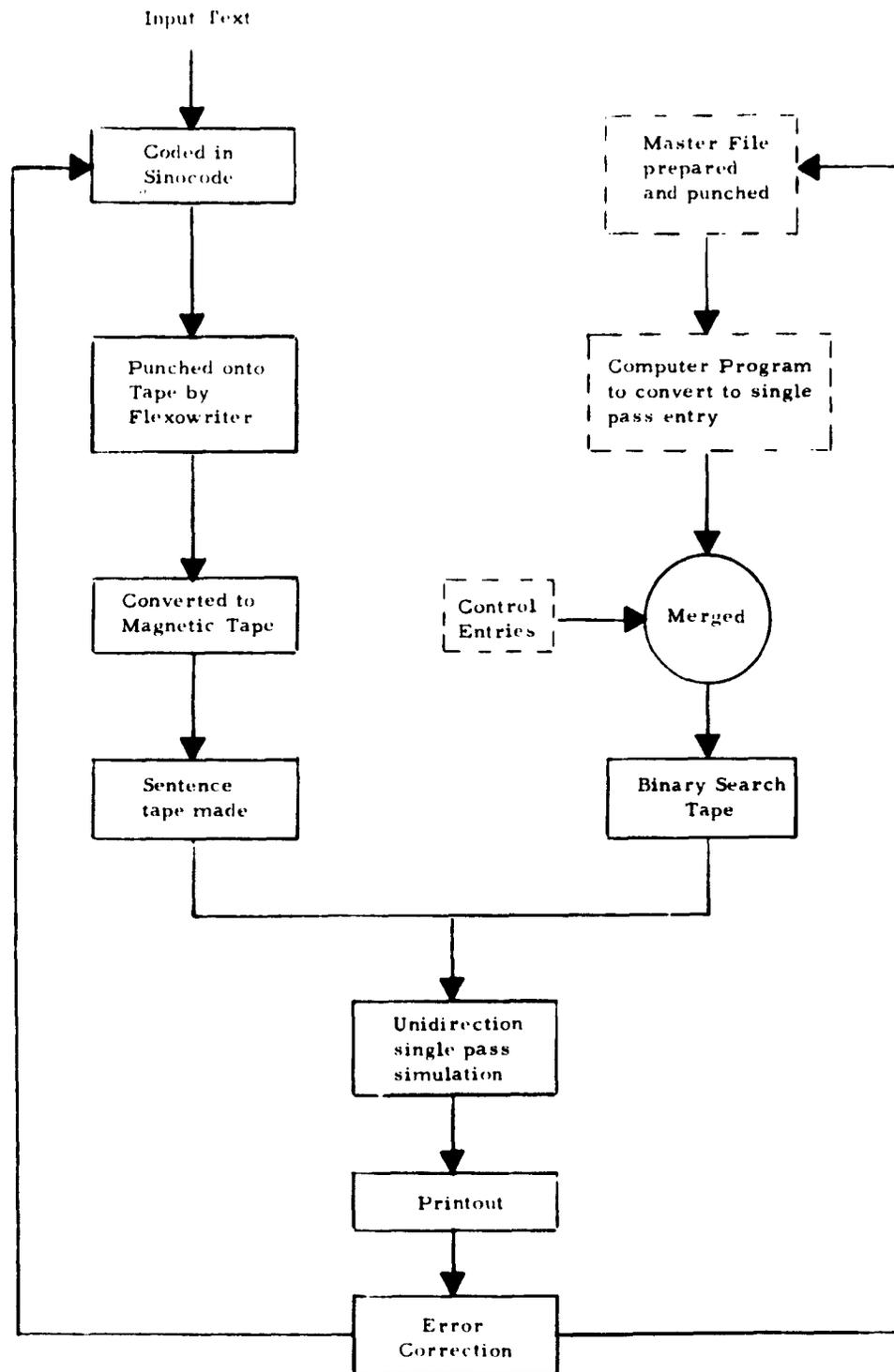


Figure 6.1 Error Detection by Unidirectional Single-Pass Translation

entries and English synthesis-pass entries, a Master Reference File is prepared and kept.

This Master Reference File consists of standard 80-column IBM cards divided into the following fields.

Field One

Column One

This field contains a single alphabetical code to denote the article from which the entry was initially derived.

Field Two

Columns 2-7

This field contains the acquisition number, necessary in all card files for housekeeping purposes. This acquisition number is the same as the acquisition number used as identification in other process words.

Field Three

Columns 9-28

Pinyin romanizations of the characters of the entry are stored in the field.

Field Four

Columns 30-50

Sinocode representations of the characters are entered in this field.

Field Five

Columns 52-55

The parts-of-speech of the entry are stored in the field.

Field Six

Columns 59-80

This field contains the English equivalent of the Chinese entry.

When any of the fields overflow, the overflow is punched in the same field on a second card. This continuation card carries the same acquisition number as the main card with an eleven punch added in column six to denote that it is a continuation card.

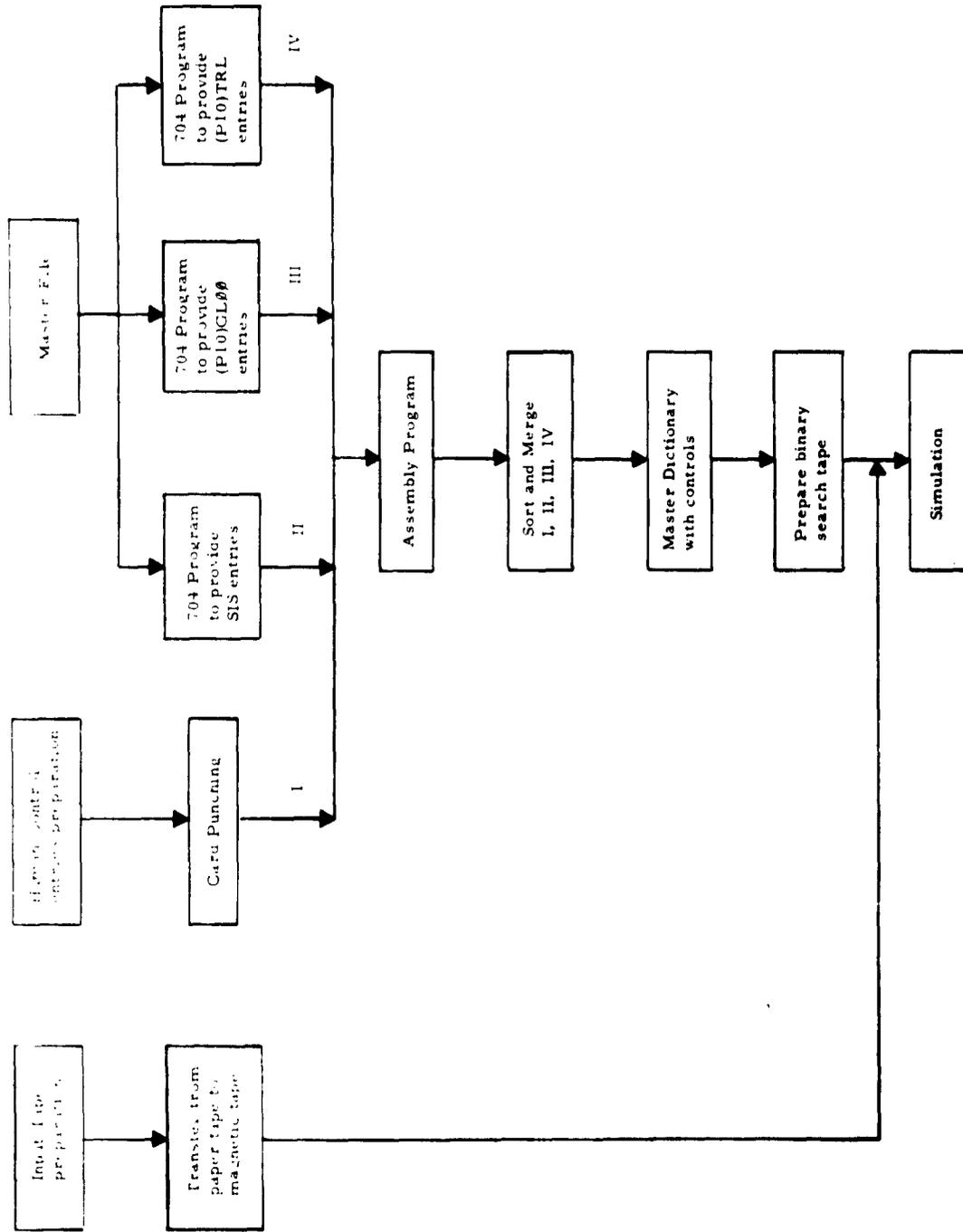


Figure 6.2 Multipass Simulation by 704 Computer

### 6.3.2 Preparation of the Master Reference File Card

To prepare the Master Reference File card, a lexicographer first selects an item as an entry in the dictionary. The Pinyin romanization part-of-speech designation and the English equivalent are also supplied. This card then goes to a Sinocode operator who supplies the Sinocode for the Chinese characters. A keypuncher enters the information in the appropriate locations on an IBM card. Acquisition numbers are serially supplied in the order the cards are punched.

### 6.3.3 Housekeeping of the Master Reference File

A 704 program was written for housekeeping purposes. This program will sort the Master Reference File in any one of the Fields, II to VI, as specified. In Field four the Sinocode field, it will also sort, as desired, on the first, second, or third Sinocode.

With this program we can derive listings of the Master Reference File in any configuration desired. Thus, if the linguist desires to know the distribution of the parts-of-speech, all he has to do is to examine the listing sorted on Field five, the part-of-speech field.

## 6.4 Generation of Multipass Entries from the Master Reference File

### 6.4.1 Search Input State Entries

The Search Input State entries have the following form:

<u>Acquisition No.</u>	<u>Argument</u>	<u>Function</u>
------------------------	-----------------	-----------------

The acquisition number is used for control purposes, the Argument allows identification of the correct entry from the input material, and the process word is the function of the argument.

The process word has the following format:

POS WOR PASS GRAMMAR ACQ. NO SEMANTIC TRAN.

This is a fixed-length word of 37 bytes divided into the following fields:

Field One Bytes 1-4

This contains part-of-speech information.

Field Two Bytes 5-7

These three bytes provide space to record the pertinent word-order information derived from analysis.

Field Three Bytes 8-17

Ten bytes are reserved for tags to carry along syntactic information of the appropriate pass during multipass analysis of the sentence.

Field Four Bytes 18-22

Grammatical information as derived from the analysis is recorded in these five bytes.

Field Five Bytes 23-28

The acquisition number of the Master File card is transferred here to allow for identification of the process word.

Field Six Bytes 29-36

Any semantic information and other pertinent structural information discovered during analysis will be stored here with appropriate coding.

Field Seven Byte 37

After a translation has been made, a tag is placed in this byte to denote the fact that this process word has been translated.

Fields one and five contain information which has to be supplied initially. All other fields store information derived during analysis.

We have automatically generated all Search Input State Entries by means of a 704 program. This program performs the following:

1. Adds 30,000 to the acquisition number in Field II of the Master Reference Card and transfers this incremented acquisition number into the acquisition number field of the Search Input State Entry.
2. Transfers Field IV of the Master File, containing the Sinocode, into the argument field of the Search Input State entry.
3. Generates the function of the Search Input State entry, i. e., the process word, from the Master File in the following manner:
  - a. Transfers Master File Field V (part-of-speech) to Process Word, Field One
  - b. Writes (DP 18)
  - c. Transfers Master File, Field Five (Acquisition number) to Process Word Field Five
  - d. Writes (DP 9)

For mechanical reasons, to conserve a fixed-length process word, the numerals 2, 8, 9 and 0, which are two-byte characters, are represented by (Q1), (Q2), (Q3), (Q4), all one-byte characters, in the process word. This allows us to produce a fixed-length process word of 37 bytes. Forward and backward length tags, and boundary indicator, will be computed and introduced by a 7090 program.

### 6.4.2 Process State Entries

Using the Master File, we generate two other types of entries, the basic translation entries [(P10)GL00] and the single pass subroutine entries [(P10)TRL].

These entries are of the form:

Accession No.	8(P10)GL00 + * (DP24)	Master File Acquisition No.		.(EA)(DP36)A(OD)English	H,.,
Accession No.	8(P10)TRL + * (DP24)	Master File Acquisition No.	(DP8)	.(EA)(DP36)A(OD)English	TRL

The accession number is generated by incrementing the Master File acquisition no. by 10, 000 for the (P10)GL00 entries and by 20, 000 for the (P10)TRL entries. The constants 8(P10)GL00 + \* (DP24) for the basic translation entries and 8(P10)TRL + \* (DP24) for the single-pass subroutine entries are written. The acquisition number in Field II of the Master File is transferred to both entries.

Then the constants ,(EA)(DP36)A(OD) for the (P10)GL00 entries and (DP8) ,(EA)(DP36)A(OD) for the (P10)TRL entries are written. The English field, field six of the Master Reference File, is transferred to both entries. The constant confix H, , , is then written for the (P10)GL00 entry and the constant confix TRL for the (P10)TRL entry.

## VII. LINGUISTIC PROCESSING

### 7.1 Parts-of-Speech

Grammatical segments may be defined in terms of "form classes," that is, a certain "class" of such segments will fit into a particular "form." These "forms" for the most part are syntactical, although some may be morphological. Since classification is based on syntactic decisions, the groups thus classified may in turn serve for sentence recognition.

To isolate general classes of grammatical segments, it is necessary to test, by trial and error, whether or not certain grammatical segments fit into certain environmental forms. The environmental forms chosen for these tests must be independent expressions. If a certain group of grammatical segments fits in one environmental form, and at the same time the same group fits one or more other environmental forms, then and only then, can that group or class be used for further study and subdivision. Subdivision of general classes can be accomplished by testing the grammatical segments in other environmental forms.

For example, we determine verbs by the syntactic analysis of the 不 ..... 了 structure. Any lexical item in free form that can appear between this construction can be identified as a verb. On the other hand, nominal expressions may be determined by the construction 这个 ..... 好 . By the examination of syntactic constructions such as these, and by the adoption of other conventional dictionary classifications, a categorization of words by parts of speech may be made.

Generally speaking, the Chinese lexical units dissected from an infinite input stream can be grouped into the following main categories:

- a. nominal types
- b. verbal types
- c. adverbial types, and
- d. grammatical function types.

Each category can again be subdivided into various subcategories.

We have employed for our analysis therefore the following scheme of parts of speech:

A. Nominal types

- |     |      |          |
|-----|------|----------|
| 1.  | NAAØ | 大家, 我們   |
| 2.  | NABØ | 范新弼, 錢学森 |
| 3.  | NAPØ | 工程師們     |
| 4.  | NASØ | 体力劳动     |
| 5.  | NLLØ | 東, 西     |
| 6.  | NLNØ | 北半球      |
| 7.  | NLPØ | 日本       |
| 8.  | NLSØ | 前面       |
| 9.  | NLXØ | 以下       |
| 10. | NLZØ | 空中       |
| 11. | NMØØ | 一些       |
| 12. | NMAØ | 公里       |
| 13. | NMBØ | 周, 件     |
| 14. | NMCØ | 大, 日     |
| 15. | NNØØ | 發动机, 星球  |
| 16. | NSØØ | 这些, 那    |
| 17. | NTPØ | 半天       |
| 18. | NTSØ | 小时       |
| 19. | NTWØ | 最後       |
| 20. | NUØØ | 九, 几     |
| 21. | NUMØ | 百, 十     |

B. Verbal Types		V $\phi\phi\phi$
1.	VBR $\phi$	来不及
2.	VC $\phi\phi$	当, 由于
3.	VCH $\phi$	去
4.	VD $\phi\phi$	上, 出
5.	VDC $\phi$	回来
6.	VDL $\phi$	上去
7.	VDZ $\phi$	在
8.	VER $\phi$	辩不到
9.	VIB $\phi$	飞, 相反
10.	VID $\phi$	到
11.	VLA $\phi$	来
12.	VO $\phi\phi$	打字
13.	VR $\phi\phi$	启动
14.	VRC $\phi$	送去
15.	VRL $\phi$	站起来
16.	VSH $\phi$	是
17.	VTAP	送
18.	VTC $\phi$	会
19.	VTD $\phi$	看
20.	VVP $\phi$	跑
21.	VVU $\phi$	有

C. Adverbial Types		(Q $\phi\phi\phi$ )
1.	QA1 $\phi$	很
2.	QA2 $\phi$	没, 可靠地
3.	QCC $\phi$	和, 而
4.	QM $\phi\phi$	如果
5.	QU $\phi\phi$	一共

D. Grammatical Function Types (F $\emptyset\emptyset\emptyset$ )

- |                            |    |
|----------------------------|----|
| 1. FBA $\emptyset$         | 把  |
| 2. FDE $\emptyset$         | 的  |
| 3. FDI $\emptyset$         | 地  |
| 4. FFZ $\emptyset$         | 分之 |
| 5. FLE $\emptyset$         | 了  |
| 6. FLI $\emptyset$         | 高  |
| 7. FNE $\emptyset$         | 呢  |
| 8. FPN $\emptyset$         | 帶頭 |
| 9. FS $\emptyset\emptyset$ | 而已 |
| 10. FSU $\emptyset$        | 左右 |
| 11. FSY $\emptyset$        | 的話 |

Besides these four main classes, there are the following types:

- (i) Adjectives (AA $\emptyset\emptyset$ ) 高, 基本 which belongs to the Verbal Types, but which have been coded differently for convenience
- (ii) Idioms (IDI $\emptyset$ ) 打腫了臉充胖子 which generally are noun phrases and a member of the Nominal Types.

7.2 The XVN and DC Pass

Before analysis of an input sentence is begun, two preliminary and important tasks must be performed. These involve the resolution of the verb/noun ambiguity in the part of speech tag of certain words and linkage of the individual elements of a discontinuous construction.

## 7.2.1 XVN

For determination of the part-of-speech of ambiguous words in the syntax of the structure they are employed in, certain rules have been derived. The principle underlying this is that in a restricted set of situations, it is fairly certain that certain words are verbal in function. Words which do not fit into this context are arbitrarily classified as nominal. The rules are also assigned priority of employment in accordance with their reliability.

Thus, to perform this discrimination, these rules are employed in order.

1. Words one word right of the XVN word is either VDL $\emptyset$  or VDC $\emptyset$ .
2. Words one word left of the XVN word is one of the following: VTC $\emptyset$  来, 去 QA2 $\emptyset$  要 .
3. The XVN word is enclosed by 所……的.
4. If a VC $\emptyset\emptyset$  precedes the XVN word and there is no other word bearing a field V part-of-speech tag in the sentence.
5. If there is either a left or right conjunction and the XVN word is joined to a verb.
6. If a noun precedes the XVN word and there is no verb between the XVN word and the beginning of the sentence.

One rule has been derived for use in the determination of the noun.

1. If an XVN word immediately follows 的 , the XVN word is a noun.

Words that have been determined to be verbal in content are denoted in the part of speech position by  $VXN\emptyset$  and words that have been determined to be nominal in content are denoted by  $NXV\emptyset$ .

### 7.2.2 DC

At the same time that words bearing  $XVN\emptyset$  tags are looked for, a comparison is made with a list of all first elements of discontinuous constructions.

If a match is found between the word and the list, we continue down the sentence looking for the second element of the possible discontinuous construction as defined by the first element. Should such an element be found, then both parts-of-speech of the elements of the DC are rewritten, the first element being written as  $DC\emptyset A$ , the second element as  $DC\emptyset B$ .

If no such matches on the list of second elements are made, then the first element is left undisturbed and the program proceeds.

### 7.3 The 的 Pass

This pass is one of the most important for it contributes most of the information pertaining to the word order rearrangement necessary for English output.

In broad outline the pass may be divided into four subroutines, each dealing with one of the following left structures:

1. AA $\emptyset\emptyset$  的
2. N 的
3.  $DC\emptyset A - DC\emptyset B$  的
4. V 的

The main routine selects the subroutine it desires to employ by examining the word to the left of 的 . After selection, the appropriate subroutine takes over.

### 7.3.1 The Adjective 的 Subroutine

This subroutine rewrites the parts of speech of 的 as AApp and tags the MP byte of the process word of 的 to indicate that it has no translation.

### 7.3.2 The Noun 的 Subroutine

Proceeding left from the word 的 , the subroutine tags the third byte of the word-order-rearrangement bytes with E unless the word happens to be a conjunction, when it is tagged with an H in the second byte of the WOR, or it proceeds until it encounters one of the boundary markers.

The boundary markers used to denote the boundary of the left structure of the 的 clause are:

1. All V's
2. All Q's except QA's and QCC's
3. All punctuations except the semi-commas ( 、 )
4. All DC's except 为
5. All beginning-of-sentence indicators

Whenever any of these boundary markers are encountered, the left-looking subroutines stop and the right-looking subroutines begin operation.

At the same time that each word is examined to see if it is a boundary marker, it is also examined to see if it has been operated upon previously in Pass V or Pass IX. If this is the case, the subroutine stops and control goes to the right-looking subroutine.

### 7.3.3 The DC $\emptyset$ A - DC $\emptyset$ B 的 Subroutines

The subroutine denotes the left structure of the 的 by the contents enclosed by the discontinuous constituents DC $\emptyset$ A and DC $\emptyset$ B inclusive. These entries are all marked by C in the record byte of the WOR tags. After marking DC $\emptyset$ A, operation transfers to the right-looking subroutine. The prohibitions of words operated on previously apply, as in the noun 的 subroutine.

### 7.3.4 The Verb 的 Subroutine

This functions in essentially the same way as the noun 的 subroutine except that several of the boundary marks are different.

The boundary marks are:

1. All V's except VTD $\emptyset$  and VTC $\emptyset$
2. All Q's except QA and QCC
3. All punctuations except the semi-comma (、)
4. All DC's
5. All beginnings of sentences.

Process words within the structure are tagged similarly, E's in the third byte for all except H's, in the second byte for QCC's.

### 7.3.5 Right-Looking Subroutine

After the left element of the 的 word order rearrangement has been determined, it is necessary to determine the right element. We proceed as follows:

Starting right from 的, we proceed to examine for QCC's, boundary markers, and words which have been operated on previously in this pass or pass V. If none of

these are found, then the third byte of the word order rearrangement tags is marked with an E.

If a QCC is found, it is marked with H in the second byte of the WOR tags, and we continue to the right of it, repeating our previous procedures. If a boundary marker or an operated word is found, we stop.

The boundary markers recognized are:

1. All V's
2. All Q's except those preceding an AA~~ss~~ or N's
3. All punctuation, except the semi-comma ( ∙ )
4. All end-of-sentence indicators
5. All DC's
6. All beginning of sentence markers.

#### 7.4 English Synthesis

As Chinese and English sentences conveying the same semantic sense may possess different structures, it is sometimes necessary during the process of synthesizing English sentences from the process words to vary the word order. It is also necessary to introduce at times inflectional variants of the English word. These are the primary problems we face in the synthesis of English sentences.

##### 7.4.1 Word-Order Rearrangement

Let us examine the Word Order Rearrangement problems. We find that there are key words which denote the requirements of one specific family of word-order-rearrangement. Elements of the sentence which might be subject to word-order-rearrangement can be phrases or clauses. Within each of these phrases or clauses, other word-orders routines might be contained.

Presently we have resolved the WOR problem into the following word-order-rearrangement classification:

- a. 的 class, including 的, 所...的.
- b. co-verb class, including 把, 被, 向.
- c. the 比 class
- d. the 多 class

The various word ordering arrangements required are:

- a.  $(X_1 X_2 X_3 \text{的} X_4 X_5 X_6)$  CHINESE  
 $(X_4 X_5 X_6 \text{ of/which } X_1 X_2 X_3)$  ENGLISH
- b. (A Co-verb B Verb Phrase) CHINESE  
 (A Verb Phrase Co-verb B) ENGLISH
- c. (A 比 Noun Phrase Adj.) CHINESE  
 (A BE Adj. comparative THAN Noun Phrase) ENGLISH
- d. (A 多 Verb Phrase Noun Phrase) CHINESE  
 (A Verb Phrase Noun Phrase 多 ) ENGLISH

The word-ordering-rearrangement functions are allowed three bytes within the process word to record pertinent information. The three bytes are coded appropriately by the four classes of word-order-rearrangement required. They also contain information as to the position of the process word within the structure of the sentence.

Translation is carried out word for word for words not requiring word-order-rearrangement until a word which is code in the word-order-rearrangement is encountered. At such an occurrence, word-ordering subroutines, as determined by the codes, are

employed. Words are no longer translated in the order of the input sentence but, rather, in the word order of the output.

#### 7.4.2 Inflection

To handle inflectional problems, words are stored in the dictionary in both the classical canonical form, i. e., the singular for nouns, simple present for verbs, and in inflected forms, with coding in the argument to differentiate between them. During analysis of the sentence in the previous passes, words have been determined to be either inflected or not, and if so, have been coded appropriately. During synthesis, words that are coded are supplied with the appropriate inflected English equivalent, and those that are not coded are translated into the classical canonical forms.

## VIII. RECOMMENDATIONS AND CONCLUSIONS

This section of the Report will be presented in two parts: (1) Software Considerations and (2) Hardware Considerations.

### 8.1 Software Considerations

As far as software considerations are concerned, two subjects will be discussed and recommendations proposed: (a) Lexicography Sinowriter indexing scheme, Sinowriter vocabulary, Preparation of a Code Cross-Reference list, Generation of a Master Reference file; (b) Syntax: Refinement of existing linguistic analysis routines, Word-ordering routines, Concordance programs.

#### 8.1.1 Lexicography

The Sinowriter Chinese Character indexing method has been proven beyond doubt to be an efficient scheme. The fact that any person, regardless of nationality, can master the keyboard scheme is a very heartening proof of the earlier convictions of people at IBM and Mergenthaler.

At IBM Research, a total of approximately 5,000 dictionary entries has been prepared based on a manual simulation of the Sinowriter indexing scheme. A number of ambiguities of character classification has been found. Examples of these are:

形  
类  
工  
半  
与  
兄  
沸

In addition, there are also a large number of characters that were not included in the input device vocabulary. Examples of these are:

斜  
應  
則  
節  
環  
尼  
真  
零  
伽  
埭  
敝  
梭  
軌  
蜣  
螂

The solution to these problems is relatively simple. For the ambiguous characters, they should be either classified into more than one family or they should be forced into the one family that is most logical for their classifications. The advantage to the first is the elimination of possible input operator confusion and the resultant increase in coding speed. The second is the minimization of the number of dictionary entries that will have to be stored on the Photodisc in anticipation of input coding variants.

For those characters that are not included in the input device vocabulary, they should definitely be properly classified and added into the appropriate families.

One of the most basic and serious drawbacks of the present input device, however, is the inability of its handling of input characters absent in the device vocabulary. Many methods could be devised to solve this problem. One is to provide an auxiliary Stroke Sequence Keyboard. This method, on the surface, seems to offer a most straightforward solution, but on further consideration, one can see that it is not practical. The Stroke Sequence method of composing Chinese characters requires operators who are native speakers of Chinese. With this addition the Sinowriter keyboard will no longer be a true universal device. The more realistic approach will be the inclusion and proper classification of the basic Kang-Hsi 214 radicals and their simplified variants in the device character vocabulary. A definite sequence of writing or punching the radicals will have to be specified so that the operators who do not know Chinese will be able to construct the characters.

Greek alphabets and some most frequently used mathematical symbols should also be included in the vocabulary of the input device.

For the simplification of the machine translation study task, the input device vocabulary characters should definitely be provided with corresponding Pinyin codes. The inclusion of the Pinyin codes will facilitate the preparation of computer concordance programs for the grammatical study purposes of input texts.

Since the Chinese official Telegraphic code has been in existence for many years and since it is very widely used by many people in the message transmission and other related fields, it is essential that a Cross-index of character codes be compiled. This index should be compiled according to the following format:

CHARACTER    TELE-CODE    SINOCODE    MATHEWS NUMBERS  
RADICAL INDEX

This Cross-index should not only be punched on paper tapes but should also be recorded on magnetic tapes. A general-purpose computer program should also be written in order to provide the means for a high-speed and efficient conversion of codes.

To facilitate word list and MT dictionary preparation, a Master Reference File should also be compiled. The MRF should have the following format:

<u>FIELD I</u>	<u>FIELD II</u>	<u>FIELD III</u>	<u>FIELD IV</u>	<u>FIELD V</u>	<u>FIELD VI</u>
Micro-glossary identification	Unique Acquisition number	Pinyin Romanization	Sinocode	Parts of Speech	Inflectional coding
					<u>FIELD VII</u> Translation

For a detailed description of the various fields listed previously for the MRF, refer to IBM Proposal which was submitted to the Rome Air

Development Command in response to the Purchase Request No. 68-048, "Engineering Change B to AF 30(602)-2479." The proposal is entitled: A Proposal to Construct a Multipass Dictionary for the Machine Translation of Chinese.

#### 8.1.2 Syntax

As far as syntactic rules are concerned, the following is a list of the interesting items:

1. Improvement and enlargement of the machine vocabulary of discontinuous constituents.
2. Enlargement of machine analysis routines for the resolution of verb/noun parts of speech ambiguities.
3. Incorporation of machine rules to resolve the adjective/adverb parts of speech ambiguities.
4. Incorporation of machine rules for the resolution of functional/conjunctive parts of speech ambiguities.
5. Enlargement of the rules for resolution of auxiliary words.
6. Introduction of automatic machine inflection generation program.
7. Enlargement of the rules for word ordering.

The discontinuous constituents should be classified into sub-categories according to: usage, translation, syntactic function and types. New rules should definitely be included to resolve the adjective/adverb, functional/conjunctive parts of speech ambiguities of words. A well-defined parts of speech definition of adjective/adverb in particular contexts will not only facilitate the MT parsing

of sentences but will also provide a new class of clues for the correct determination and resolution of the verb/noun parts of speech ambiguities.

New rules should be added to analyze a larger number of auxiliary words. An automatic inflectional ending generation program for the English correspondents should be added also.

The most basic and also the most essential grammatical studies will be the classification and assignment of a machine translation oriented parts of speech code to input words. The parts of speech codes should be classified based on the following criteria:

1. Four to five main word categories.
2. Classification based on word forms and grammatical functions.
3. Classification based on parsing requirements.

Concordance programs should be generated to facilitate the study of input texts. The following types of concordance programs will be required:

1. Concordance based on Study words.
2. Concordance based on Discontinuous Constituents.
3. Concordance based on Parts of speech.
4. Concordance based on Sentence Structures.

## 8.2 Hardware Considerations

### 8.2.1 Control System

The system control seems to be fairly well in hand, however sufficient experience with operation has not been accumulated. Therefore

shortcomings of the system have not been developed to a point that they can be evaluated.

#### 8.2.2 Illuminants

Resolution of the system seems to be reasonable over the field presently used (26 x 26), but the illumination system is completely inadequate. Further investigation of this problem should be made in order to determine if there is a way to obtain necessary illumination over the entire field.

#### 8.2.3 Wedge System

The wedge system binary should be reduced one level (from five wedges to four) per coordinate -- if no solution to the illumination problem is developed. The resultant system would require the use of more memory plates for the same total capacity.

#### 8.2.4 Coding

The coding of the system used for scan purposes should be the master determinant rather than the Translator addresses. This might be impossible to accomplish in practice but a separate coding consideration will definitely simplify the overall circuitry requirements of the viewer and the punching system.

#### 8.2.5 Electronics

The stepper control circuitry is cumbersome and should be redesigned.

A stepping relay should be substituted for the input and output control relays.

The electromechanical linkage system for the prism drive mechanism should be redesigned for ease of service and accessibility.

The memory plate changing mechanism should be replaced with a more sophisticated design. The drum device was designed simply to indicate an approach to coded control of plate selection, and excessively restricts the optical design.

#### 8.2.6 Photographic Plates

These should be prepared in separate quadrants to allow easier alignment of the system. It will aid in matching the wedges and will also allow larger tolerances for the plates.

**APPENDIX I**  
**The Sinowriter**

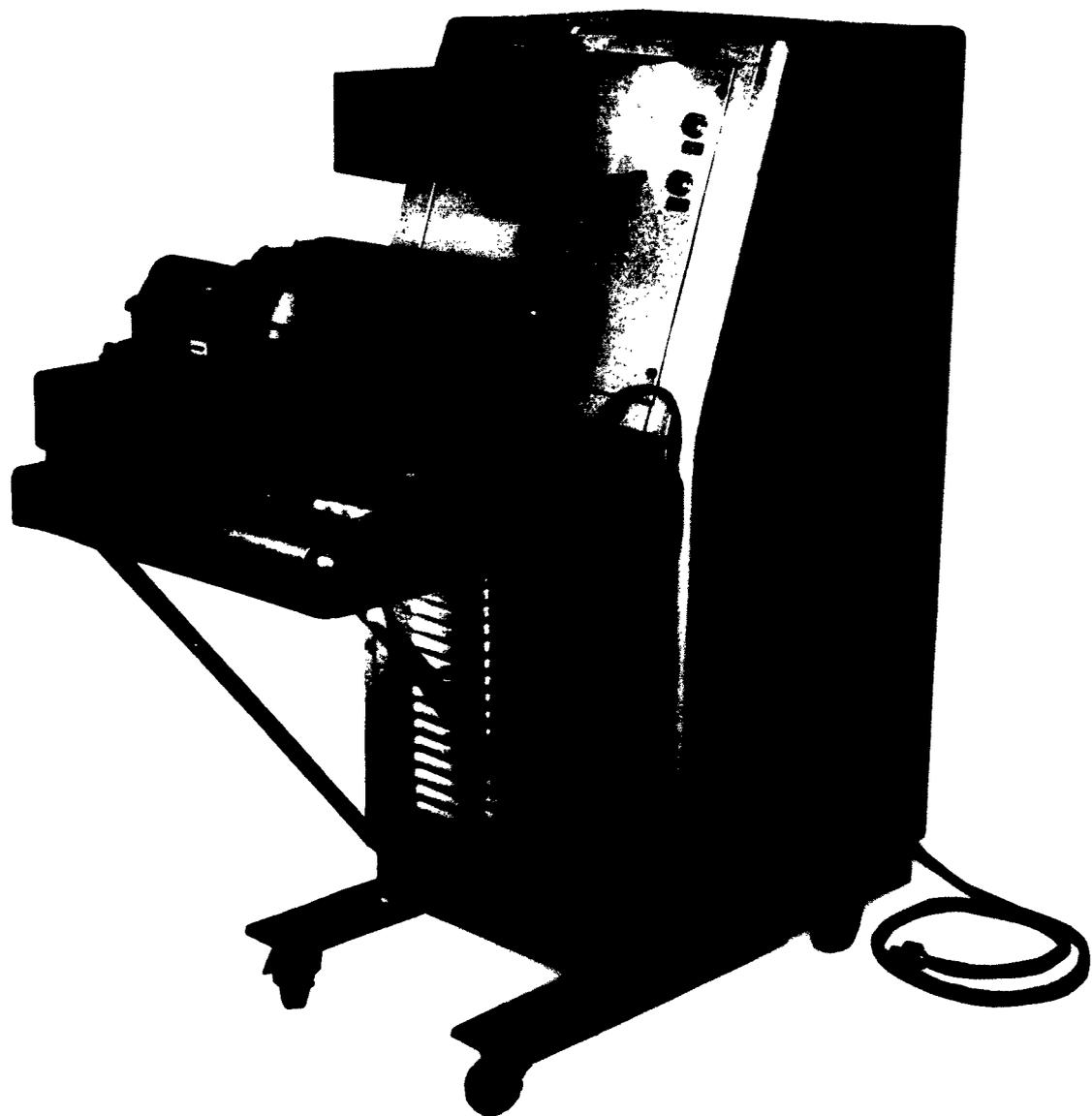


Figure A1.1 Chinese Input Device

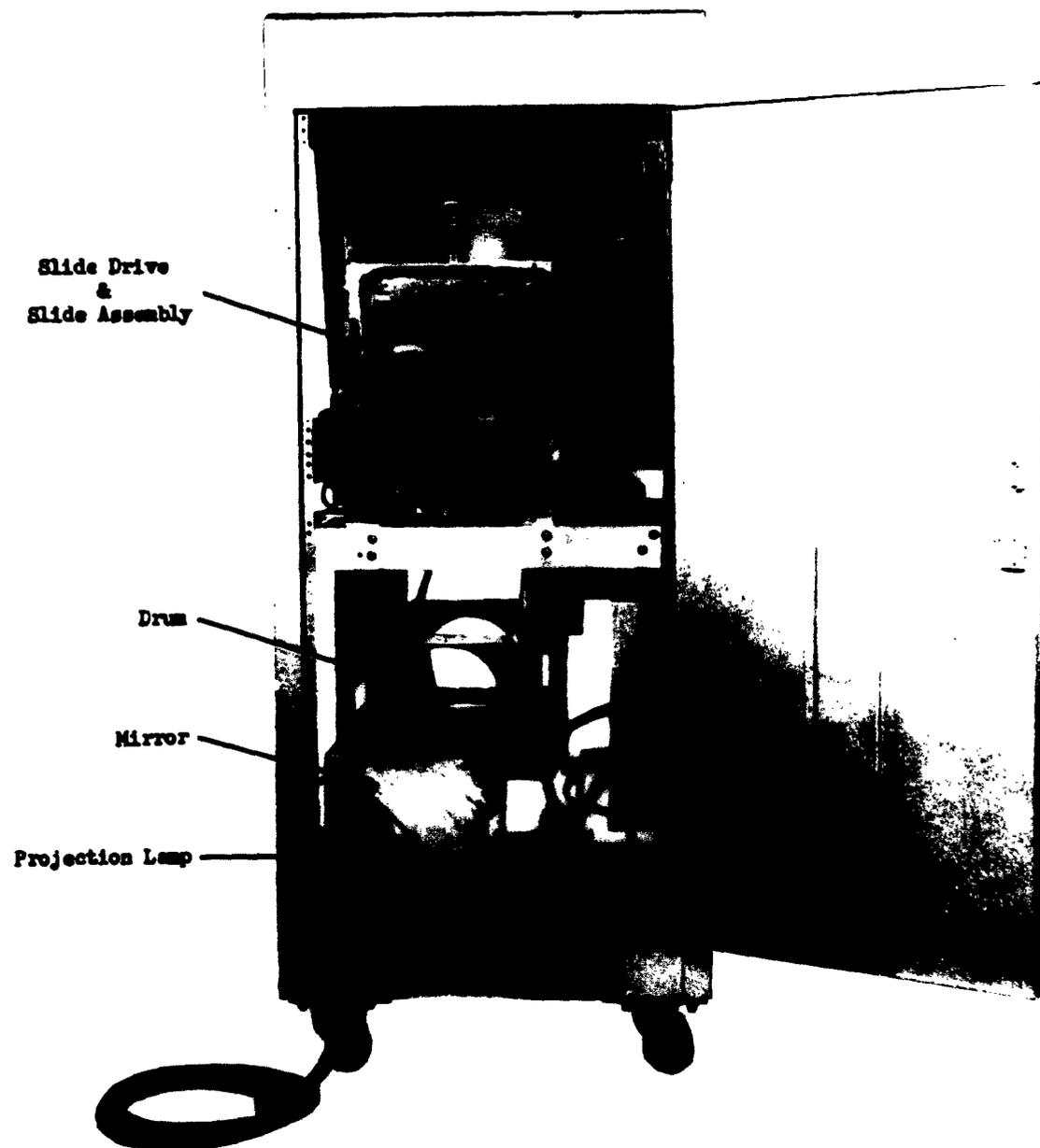


Figure A1.2 Chinese Input Device, Rear View

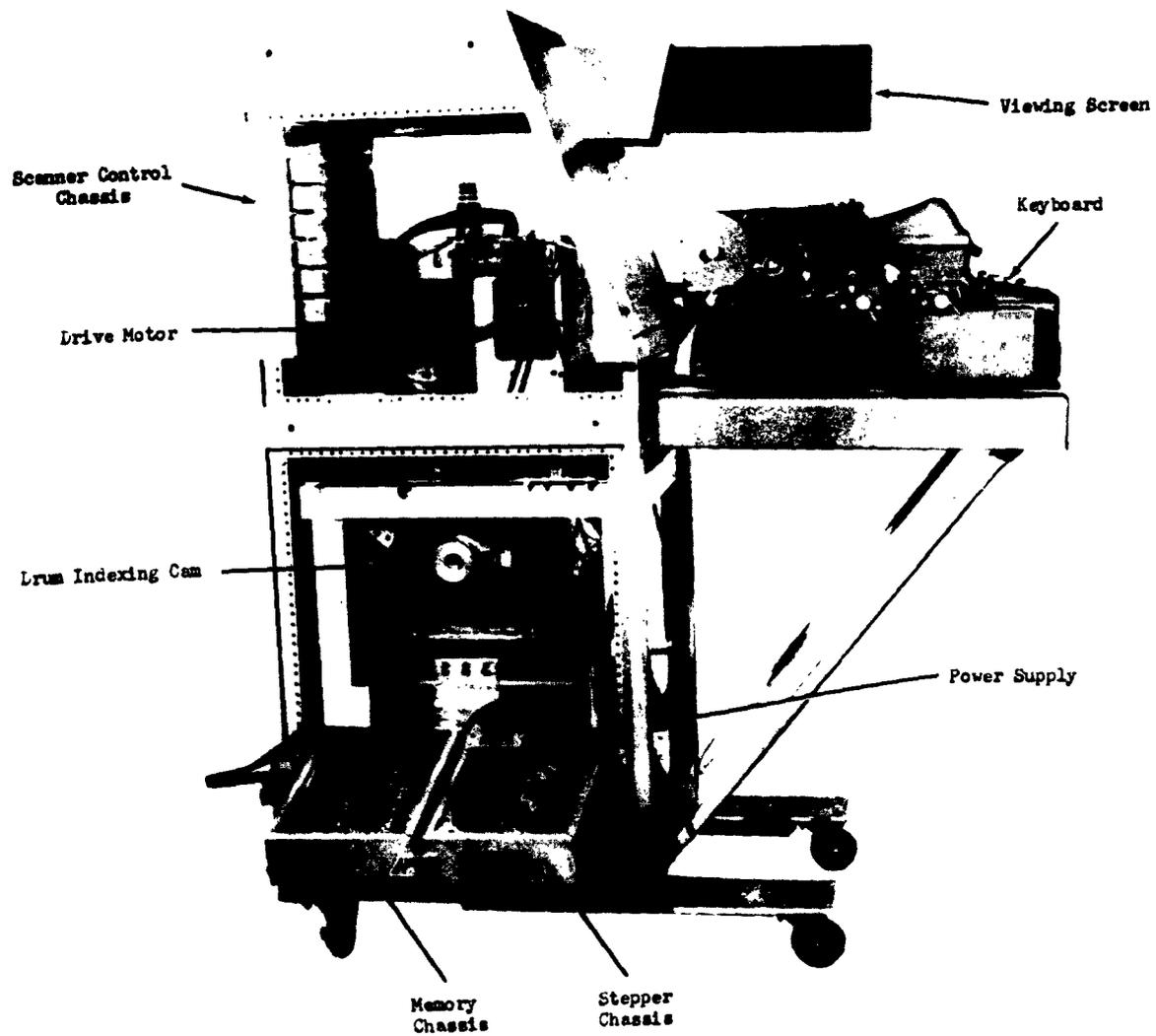


Figure A1.3 Chinese Input Device, Side View

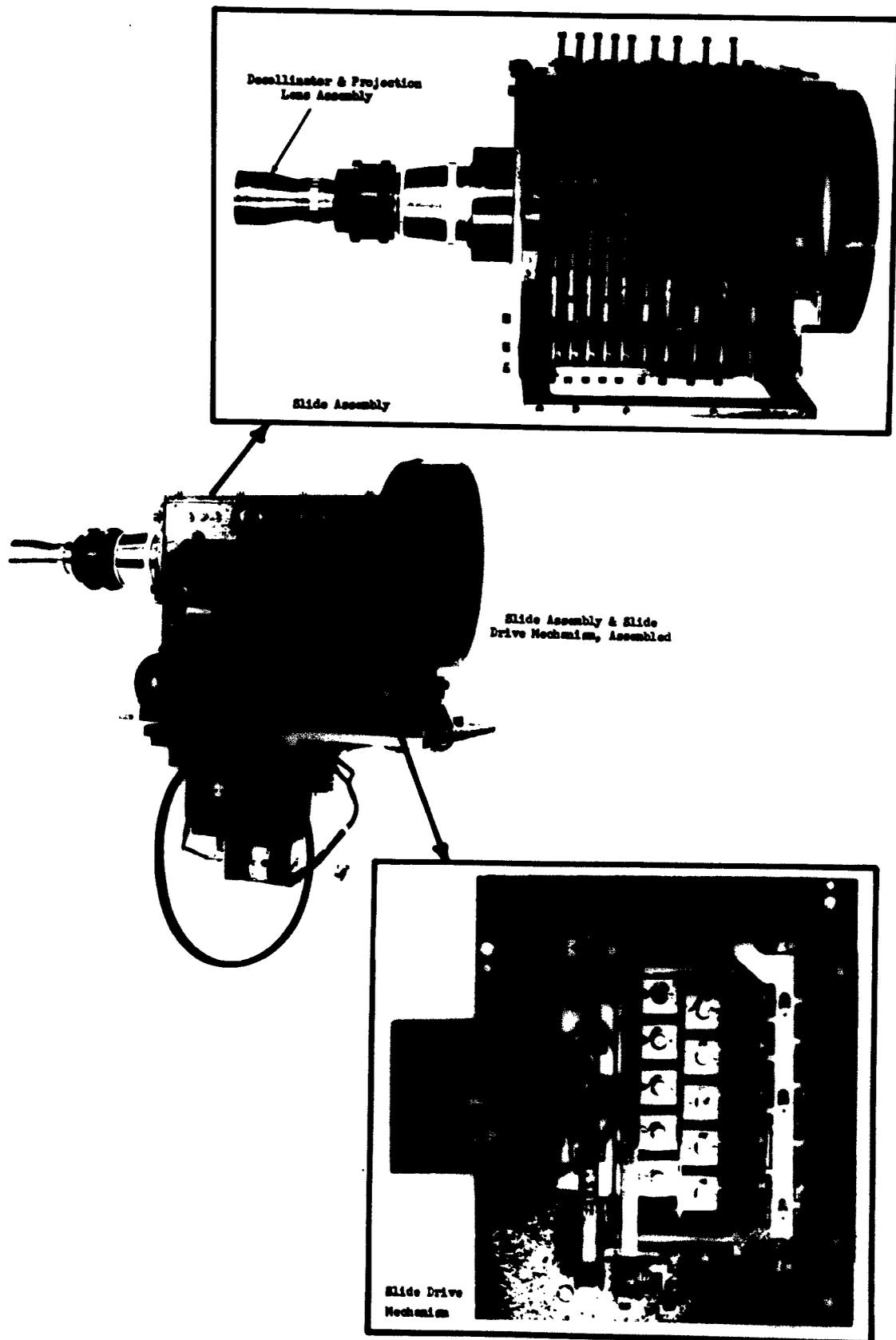


Figure A1.4 Slide Drive & Slide Assembly

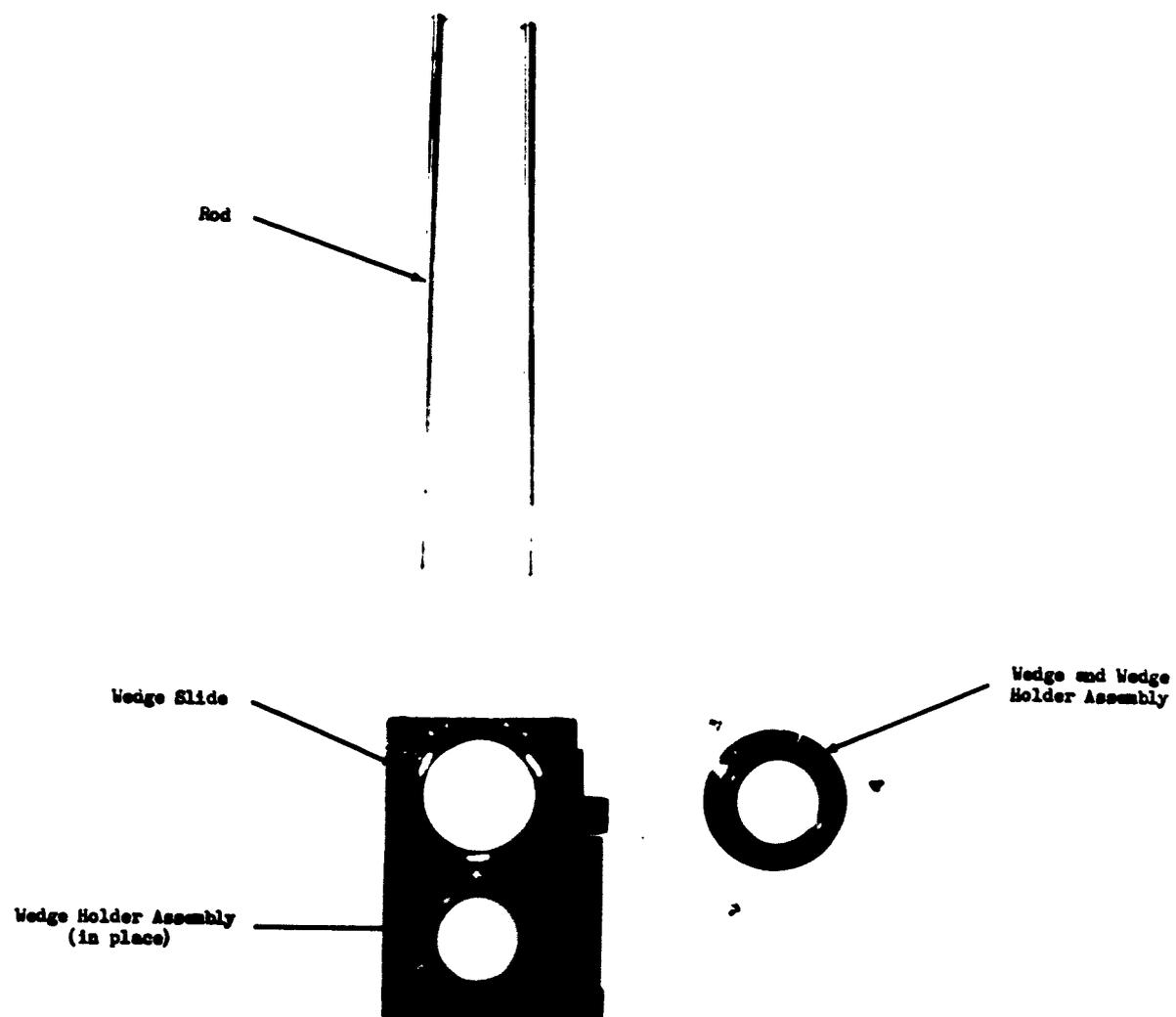


Figure A1.5 Wedge Holder Slide

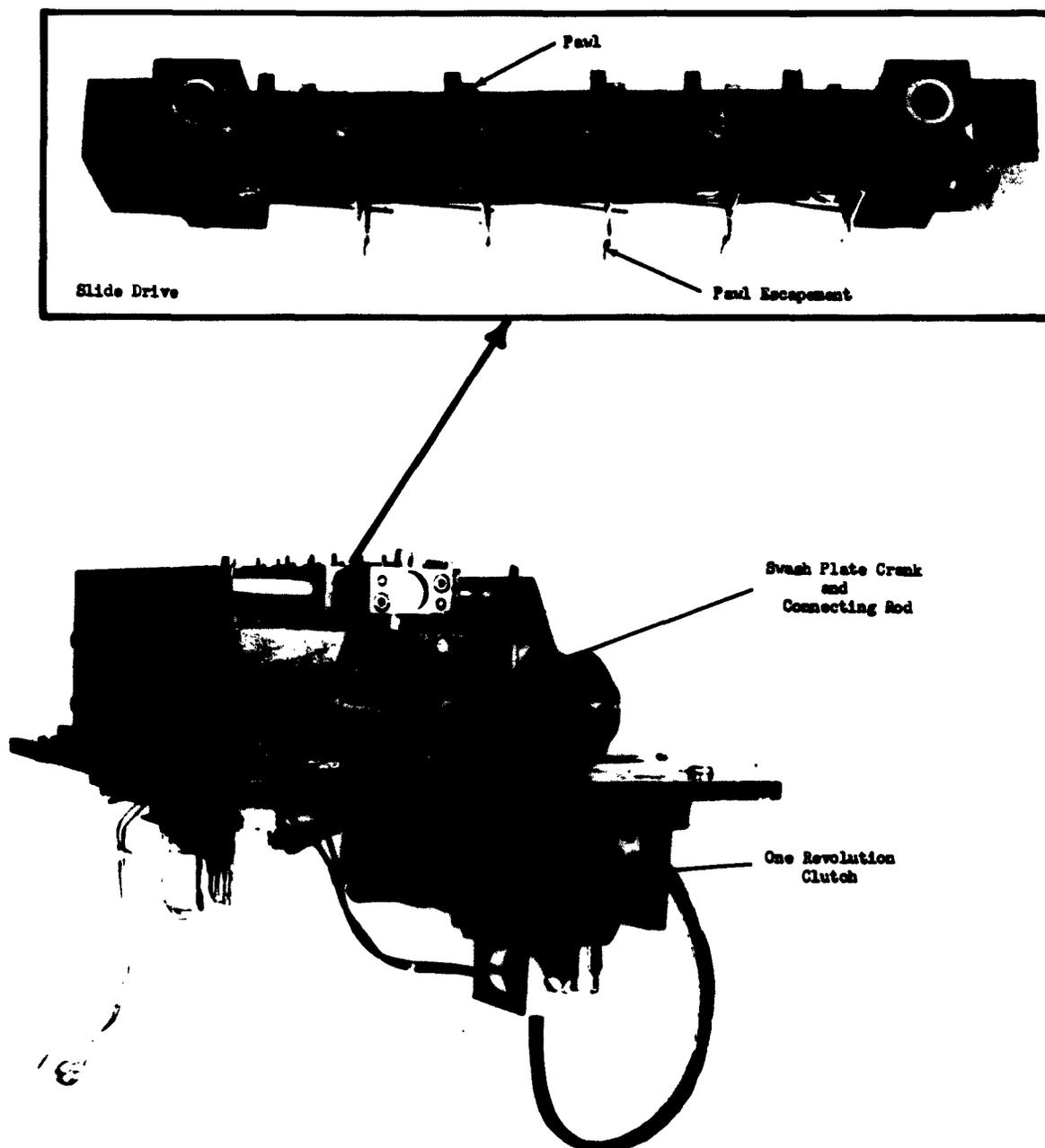


Figure A1.6 Slide Drive Mechanism

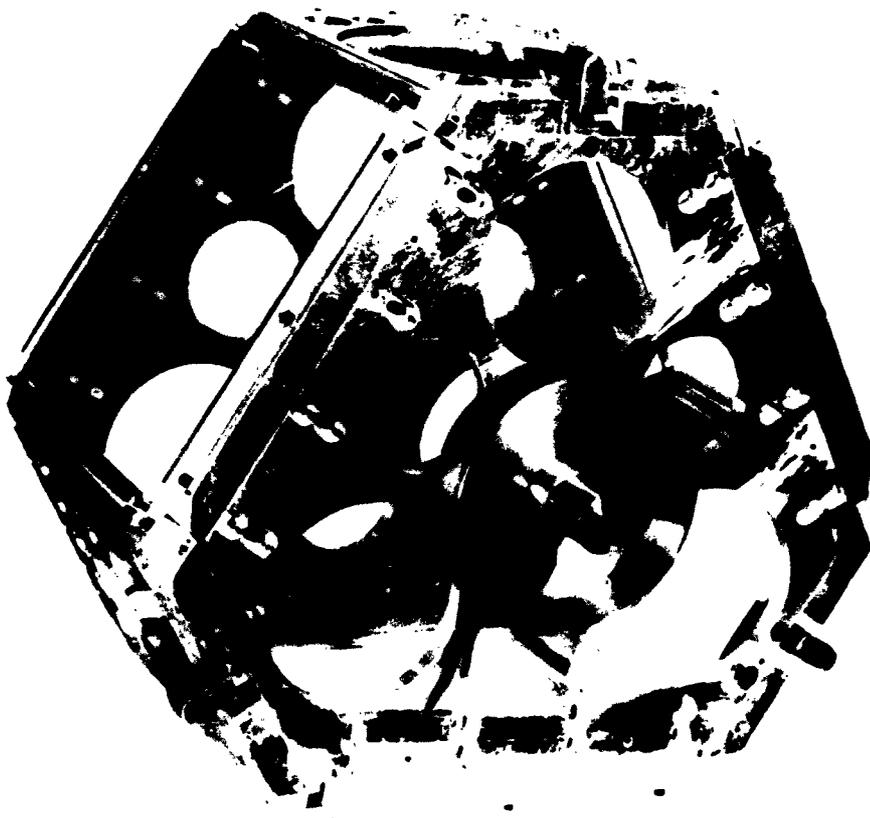


Figure A1.7 Drum

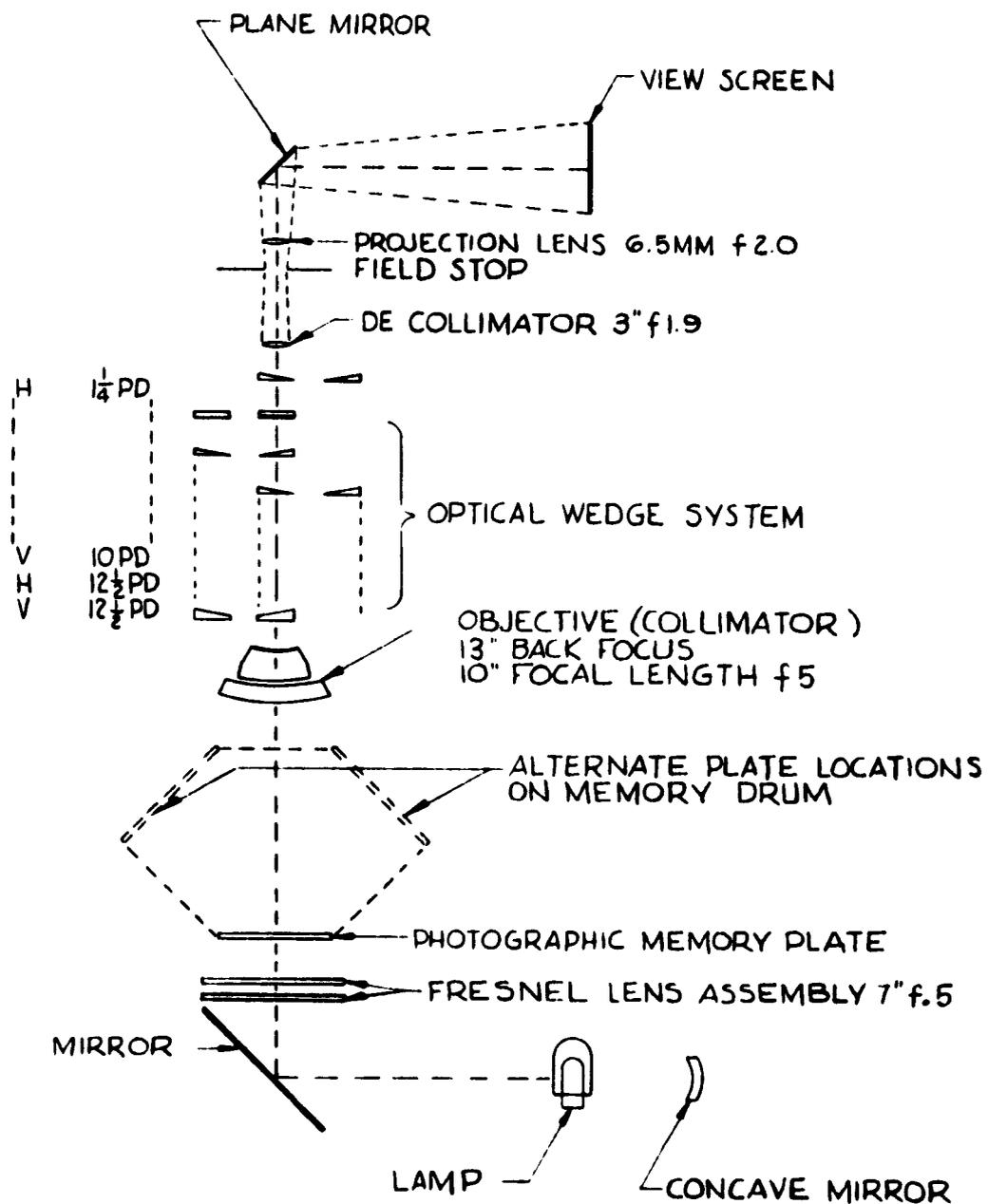


Figure A1.8 Optical Schematic Chinese Input Device

**APPENDIX II**

**Samples of Chinese-English Machine Translation Output**

## 軍事常識叢書

# 防 空 軍

中國青年出版社

### 防 空 導 彈

在現代防空軍的隊伍里，又增加了一種新的部隊，這就是防空導彈部隊。在第二次大戰後期，英國的首都倫敦和其它地區，曾經被德國的“無人駕駛的飛機”襲擊過。這種當時被人們叫做“無人駕駛的飛機”，實際上就是導彈。不過，那時候的導彈在技術上還不夠完善，命中率很小。

導彈受地面或空中無線電的操縱，在無線電的引導下飛行。它可以作為進攻的兵器，也可以作為防禦的兵器。近幾年來，世界各國都很重視導彈的研究和生產。英帝國主義為了準備新的侵略戰爭，正在瘋狂地製造各種類型的導彈，從1946年以來，研究導彈的費用就花了六十多億美元。現在的導彈已能帶上原子彈和氫彈的彈頭，因而它是一種破壞力很大的兵器。

現代防空軍裝備的防空導彈，是一種最新式的防空兵器。它不但可以對付一般的空襲兵器（如飛機），而且可以有效地打擊敵人放出的導彈。防空導彈主要分兩種，一種是地對空導彈，另一種是空對空導彈。

Translation taken from

Military Knowledge Series

**Air-Defensive Force**

Peking

**Air-Defensive Guided-Missile**

Page 25

Within ranks of modern air-defensive corps ,also has increased one kind new army corps,this really is air-defensive missile army corps. Toward the end of the Second World War ,British capital London and other/others areas,already have been attacked by Germany pilotless aircraft". This kind at that time called by people"pilotless aircraft" in reality really is guided missile. But, guided-missile of that time upon/on skill still not perfect,chance of hitting target very small. Guided missile controlled by ground or air radio,under guidance of radio flies. It capable of used as offense weapon,also capable of used as defense weapon. In the last few year ,world every nation all very pays attention to manufacture and research of guided missile. American imperialist for the sake of preparing new aggression warfare,at the moment madly fabricates guided-missile of every kind model,ever since 1946year ,expenses of research guided-missile already has spent more than sixty billion American Dollars. Modern guided missile already possible carry with war head of hydrogen bomb and atomic bomb,therefore it is one kind weapon with very big power of destruction. Air-defensive missile that modern air-defensive corps equip is one kind most modernistic air-defense weapon. It not only capable of deals with generally air-raid weapon(like airplane)but also capable of effectively attacks enemy shoots out guided missile. Air-defensive missile mainly divides two kind: one kind is ground-to-air guided missile,another one kind is air-to-air guided missile.

# 数字计算机的发展方向

范新弼

计算技术是现代科学中能起带头作用的一门综合性科学。它能迅速而且准确地解决科学研究、国民经济和国防上非常复杂而繁重的计算问题，也同样能在这几方面直接地参加操作而起着自动控制和自动调节的功能。应用计算技术，能大大地减轻人类的体力劳动和脑力劳动，同时也就大大地提高了生产率。

目前计算机主要应用在下列三个方面：

(1) 解决科学研究、工程设计和国民经济的复杂而繁重的计算问题；(2) 结合运筹学，在国民经济上作经济分析、会计统计和计划安排；(3) 在自动控制系统上作起中心枢纽作用的控制机。

除这几方面外，还可以用计算机来研究信息加工、控制论和逻辑问题，如文字翻译、论文摘要、机器图识别、下棋、战术研究以及模拟和研究人的感觉和活动过程等（当然这种研究并不是将脑的活动完全归结为无机物的活动，而只是利用计算机来对脑的部分活动作较深入的了解），研究的结果反馈来也会对计算机的逻辑设计提供新的方案。

由于计算机的应用不同，所以就有各种不同类型的计算机，如通用机和专用机，以及大型、中型、小型等。计算机的发展非常迅速，现在世界上已有了近万台计算机。科学研究和国民经济上复杂而繁重的问题多半是在大型通用计算机上来解决，因为它需要快速地进行大量的运算，而在其他方面的应用上，各种类型的计算机都有。但总的说来，从性能上对计算机的要求可大致归纳为下列几项：

(1) 可靠性。这是计算机最基本的要求。计算出来的结果如不可靠是没有用的。苏联成功地发射人造宇宙卫星而美帝国主义却多次失败，原因之一就在于可靠性方面。据估计美国的自动控制和计算设备在高振动高温下失灵，而使火箭脱离了轨道。为了使结果可靠，首先是不要错，其次是万一错了也要能检查出来并自动纠正。要机器不要算错就要求机器的各个部件、线路和元件都可靠地工作。为了保证机器能正

常工作，除上述两点外，还要使机器少发生故障，易于检查、修理和维护方便。但是这几乎是不可能的，因为元件和部件有老化变质和寿命的问题，线路上有干扰现象，加上电源和工作环境不稳定都会影响到它的工作可靠性。一般来说，真空管和继电器是不太可靠的元件，所以将来的计算机采用晶体管或其他固体元件是肯定的方向。由于元件的不完全可靠和性能不稳定，用过多的元件来做计算机是不好的，但如元件太少而使操作不完善，也是不可靠的，所以在线路设计时，就应该考虑到元件的数值在适当的范围内变化，而线路仍旧是可靠地操作。但由于线路难以完全可靠，就要在考虑计算机的线路设计和逻辑设计时采用优良的系统及适当地加入检查系统使能避免出错。当然适当的使用和维护都会增加计算结果的可靠性。总之，要保证计算机在一定时间内计算结果的可靠是必需的。

(2) 快速。这是目前对计算机的发展上提得较多的一种指标。当科学研究和国民经济分析的问题越来越复杂，数据愈来愈多和自动控制的反应时间要求愈来愈短时，都需要有更快的机器操作速度。在原子能和力学（人造卫星）的计算上就有百万次到亿次运算操作的问题，速度不快的计算机就不能满足要求。所以更快速的计算机是发展的一个方向。要内部运算速度快，主要就是要使运算器和内存器的操作快。运算器的问题在于单元线路的快速操作和讯号的迅速传递。因此除要元件能高速工作外，还需要采用新的逻辑线路来解决。据目前的资料，已有运算器线路用到50兆周的脉冲基波。内存器一般的容量是4096个数（最大到六万个数），存取时间4—6微秒，也有容量较小而存取时间仅0.5微秒的存储器，但这种线路在应用上还存在一定的問題。

目前电子管线路已快接近极限，所以提高操作速度应从新元件（如晶体管、铁磁材料等）和新线路（如动态线路）着手。再者，现有的脉冲技术应用到快速线路上还有问题，比如高频机脉冲的产生、放大和传递时出现的驻波、辐射、干扰和损耗等。所以线路

开拓一门新的孔微形脉冲技术来解决这些问题。

内部运算的问题解决后，还没有完全解决快速的问题。适应快速计算机的输入输出设备和如何最合理地使用各个部件和自动化编制程序等问题也是需要解决的。

(3) 存储量。在科学计算、统计、控制和訊息加工、翻译等工作中要求有极大的存储容量，但下一定要很快的操作速度。

现有的水平是：

容量(二进制)	存取时间 (包括等待时间)
磁鼓设备 $1.7 \times 10^6$	0.2秒
磁带设备 $2.8 \times 10^8$	300秒
磁盘设备 $4 \times 10^7$	0.7秒

光电存储和磁鼓的容量差不多。

在国民经济分析中数据处理上已要求有存储量从百亿到一万亿二进制。距目前亿位存储量要多出百倍到万倍。

要提高存储容量，除继续改进上述设备外还需要寻找新的存储方法和新的存储介质材料。

(4) 灵活性。主要求计算机性能良好和能解决多种类型的问题外，还要求计算机有扩充的余地；在增加几个部件后，操作和算题的能力可以大为增加。这样，计算机的结构可能包括很多部件而成为一个计算机系统。另外，在使用上也可以把计算机当作一个部件来使用；几个机器并行使用和把计算机与各国的通讯网结合起来，可以把几个在不同地方的计算机联合起来进行运算，或把各地的情报数据直接送到计算机来操作，这就可以大大地简化和加速计算机的使用。

(5) 经济。从投资费、维护费、功率消耗、体积和重量来看都要求经济。

## 二

为了满足上面这些要求，计算机需要进行研究下列几个方面：

(1) 计算机总体系统的研究。十几年来虽然做了很多计算机，但它的总体设计还没有完整的理论基础。新近大型快速计算机已经不只有一个运算器、存储器、输入输出设备，而是有几个运算器、存储器、输入输出设备同时工作，事实上已经是一个计算机系统。什么样内部结构才好还有待于深入的研究。对于一定大小的计算机而言，如何使它的结构更灵活和有扩充的余地，以至于能解决更多种类型问题？数字机与模拟机可否结合和如何结合？元件是不可解绝对

可靠内，可否用不可靠的元件部件做成可靠的机器？如几个计算机同做一题或一个机器同做几题求校对检查，或者机器能自动发现错误且能自动纠正。此外，元件部件的工作频率在某一个时间有一定的限制，但计算的速度可否加快，和它的效率可否提高（如几个机器结合起来同做一个题目，或一个机器采取分阶段同时工作下题目）？这些都需要数理逻辑工作者、数学工作者和工程师们共同研究解决的问题。这将是理论性工作，当然要好好结合实践来进行。另外，心理学上用计算机来探索人脑的记忆和思维过程，这种研究的结果也会有助于计算机逻辑结构的设计（例如照像机、雷达和控制论，都是从对人体机能的模拟而创造出来的）。

(2) 程序自动化。目前机器是自动化了，但是算题的时间还很长，原因是在于要人编制程序后，再输入机器去操作。这些数据和程序经过人的加工，又难免不发生错误。目前这种错误还是很大的问题，所以为了使算题的过程可靠并且快速，机器自动编程序是很重要的。这是目前程序工作者大力研究的一个问题。

(3) 元件和部件的研究。要使计算机满足上述要求，除从逻辑结构和程序方面致力外，还要从机器本身及其单元入手，才能解决问题。所以新元件和部件的研究是很重要的。计算技术要用到两种无源电元件，即有源元件和无源元件。有源元件中，从工作速度、体积、重量、消耗功率、寿命及可靠性等来谈，电子管都不如晶体管。现在已有计算机的晶体管电路在工作 26,000 小时后来发现衰老现象。与电子管几百、几千小时的寿命相比较，晶体管电路的可靠性是大大地提高了。所以晶体管是肯定要发展的方向。晶体管在频率、功率性能和温度特性等方面，目前还不能完全满足要求，需要继续研究和改进。无源元件包括电阻、电容、脉冲变压器、延迟线、存储元件、逻辑元件、接头等。从可靠性来看，要求元件在可能的使用情况（高温、高压、强震动、辐射等）和时间内操作可靠，数值稳定。从快速和功率的观点来看，元件的小型化是很重要的，而尤其重要的是新型元件的研究。用铁磁材料、铁电材料、超导体材料等组成的固体元件电路是新的发展方向。目前用铁磁材料做记忆元件有非无源器件的优点，几乎所有新研制的计算机中，都用磁心元件做快速存储器。过去以为磁心元件的开关时间限制在微秒左右，这样，用它做成的存储器周期就限制在几微秒。近来发现磁心的开关时间可以缩短，做用它来做更高速的存储器了。当然，多孔磁

心、磁通、磁导等新元件还是一个主要的研究方向。用磁心做逻辑元件，尤其是变参数元件是值得研究的。

铁电元件是用电场来操作的，所以它是一种电压型元件。在这一点上它是和真空管同一类别的。也就是说把它们用在一起较适合。铁磁和超导体元件是用电流产生磁场来操作，所以它们是电流型元件，这是和晶体管相似的。这样，用大功率晶体管来直接推动磁心是较好的配合。超导体元件的优点在于它体积小和消耗功率小。但铁电体和超导体元件尚受到一定的限制，不能大量的应用。如果能够找到一种性能（电滞回线矩形性、温度、频率等）好而制造容易的铁电材料，或者在常温（或不太低的温度下）有超导现象的材料，它们都会迅速地能在计算机上被采用。目前它们的研究工作还多半在材料的性能方面。

Development direction of digital computer.

Fan Hsin-1 i.

Computation technique is modern age science possible one composite science of produce leading effect. It possible fast but also accurately solve very complex but tedious computation problem of science research, national economy and national defense, also similarly possible at this several direction directly participate operate but effected function of automatically regulation and automatically control. Application compute technique, possible greatly reduce mental labor and physical labor of human, at the same time then also greatly has elevated production rate.

One.

Present time computer principally application at below mentioned three direction.

(1) solve tedious but complex computation problem of national economy and science research, engineering design;  
 (2) combination operations research, on national economy make economy analysis, accounting statistics and plan arrangement;  
 (3) on automatically control system make controller of produce center key effect.

side from this several aspect, still possible use computer in order to study data processing, control theory and logic problem, such as activity process and sense of research human and simulation and language translation, abstract of thesis, machine library, chess playing, and so on (of course this kind research is definitely not activity of brain completely included as activity of non-living substances, but is only

use computer to make towards part activity of brain comparatively penetrating understanding), result of research turned around also possible propose towards logic design of computer new scheme.

Because application of computer dissimilar, therefore then has computer of various kind dissimilar type, if/ such as general-purpose machine and special purpose machine, and large type, middle type, compact and so on. Development of computer very fast, presently world already with near ten thousand set computer. Science study and national economy complex but tedious problem mostly is on large type general-purpose computer to solve, because it require quickly proceed with great quantity processes, but, application of at other aspect various computer of kind type all has. But taken as a whole, from property may towards requirement of computer generally classify as below mentioned several item.

(1) reliability. This is computer most basic requirement. Result which has been calculated if/ such as not reliable is that which have not use. Soviet Union successfully launch artificial cosmos stellar body but American Imperialism then many times failure, one of the reasons then is due to reliability aspect. According to that of estimate United States automatically control and compute equipment under high vibration high temperature fail, but cause rocket detached from trajectory. For the sake of cause result reliable, firstly is not if/ want wrong, secondly is in case has wronged also if/ want possible has been found also automatically correct. If machine not if calculate wrongly then component and various part, circuit of requirement machine all reliably work. For the sake of guarantee machine possible normally work, aside from

above mentioned two, still if/want cause machine give rise to less obstruction, easy to examine, repair and maintain convenient. But this almost is not possible, because component and part has problem of life and deterioration due to aging, circuit has interference phenomena, add to power source and work environment not stable all possible has affected its work reliability. Generally speaking, vacuum tube and capacitor is not reliable component, therefore computer of future utilize transistor or other solid-state component is has determined direction. Because that of component not completely reliable and property not stable, use exceedingly much component in order to make computer is not good, but if/such as component few/less but cause operation not perfect, also is not reliable, therefore at the time of circuit design, then should considered value of component in appropriate scope variation, but circuit still is reliably operate. But because Secondary prefix error final pass Of course appropriate utilize and maintain all possible increase reliability of computation result. All in all, if/want guarantee computer in definite time reliable of computation result is necessary.

(2) high speed. This is present time mentioned towards development of computer comparatively more one kind indicator. At the time when problem of national economy analysis and science study more and more complex, reaction time requirement more and more short which automatically control and data more and more, all require has even fast/almost machine operation speed. On atomic energy and mechanics (artificial stellar body) computation then has million time problem of hundred hundred thousand time arithmetic operation, speed not fast/almost computer then not possible satisfy requirement. Therefore computer

of even higher speed is one direction of development. If/want interior processes speed fast/almost, principally is just if/want cause operation of interior storage unit and arithmetic unit fast/almost. Problem of arithmetic unit is due to fast transmission of signal and high speed operation of building block circuit. Therefore aside from if/want component possible high speed work, still require use new logic circuit to solve. According to material of present time, already has arithmetic unit circuit used for clock pulse of 50 megacycles. Interior storage unit generally capacity is 4096 digits (most large six ten thousand digits) access time 4--6 microsecond, also has storage device of access time only .5 microsecond but capacity comparatively small, but this kind circuit on application still existing definite problem.

Present time electron tube circuit already fast/almost approaching class limit, therefore elevate operation speed should from new component (such as transistor, ferromagnetic material and so on) and new circuit (if/such as transient circuit) begin. Furthermore, presently available pulse technique applied to high speed circuit still has problem, such as loss and standing wave, radiation, interference which transmission time appear and product, amplification of high frequency narrow width pulse and so on. Therefore then if/want pioneer one new milli-micro second pulse technique to solve these problem.

Problem solution after which interior processes, still have not completely solve high speed problem. Suited to how and input-output equipment of high speed computer fastest and best utilize various part and automation program and so on problem also is require that of solution.

(3) storage, capacity. In science compute, statistics, control and signal process, translation and so on work requirement has extremely large storage capacity, but not definite if/ want very quickly operation speed.

Presently available level is.

Photo-electricity store about the same as capacity of magnetic disk.

On national economy analyse and data process already requires has storage capacity from hundred hundred thousand to one ten billion bits. From present time hundred thousand position storage capacity if/want exceed hundred times ten thousand times.

If/want elevate storage capacity, aside from continuously improve above mentioned equipment still require search new storage method and new storage medium.

(4) flexibility. Aside from requirement computer property good and possible solve many different types problem, still requires computer has room for that of expansion; after increase several part, ability of computation and operation possible is greatly increase. Thus, structure of computer possible include

very much/more/many part but become one computer system. Besides, on utilize also possible computer taken as one part to utilize; several machine in parallel utilize and coupled together communication net of entire country and/with computer, possible computer of several at the place of dissimilar joined together proceed processes, or deliver to information data of various places directly computer to operate, this then possible greatly simplify and accelerate utilize of computer.

(5) economy. From the viewpoint of investment cost, maintenance cost, efficiency consume, dimension and weight all requires economy.

Two.

For the sake of satisfy above these requirement, computer require proceed research below mentioned several aspect.

(1) research of computer logic system. In the last ten plus years although has made very much/more/many computer, yet its logic design still have not complete theory basis. Recent large type high speed computer already not only has one arithmetic unit, storage device, output equipment, but is has several arithmetic unit, storage device, input and output equipment at the same time work, in reality already is one computer system. Logic structure of what kind the best still has to await penetrating research. Talk about computer of definite size, if/such as cause towards its structure even flexible and has room for that of expansion, so as to possible solve even more type problem. Digital computer and/with analog machine may or may not combine and if/such as combine towards. Component is not possible absolutely

reliable, may or may not use not reliable component part accomplish reliable machine? If/such as several computer together work one topic or one machine together work several times to verify examine, or machine possible automatically discover error and also possible automatically correct. Besides, work frequency of component part in the certain one period had definite limit, but speed of computation may or may not accelerate, and its utilize rate may or may not elevate if/such as several machine combined together together work one topic, or one machine adapt multi-programming techniques at the same time make several topic? These all require mathematical logic workers, mathematics workers and engineers together study problem of solution. These is theory work, of course if/want suitably combine practice to proceed. Besides, psychology use computer in order to explore thought process and memory of human brain, result of this kind research also possible be of assistance to design of computer logic structure (for example camera, radar and control theory, all is from but created towards simulation of human body mechanism.

(2) program automate. Present time machine is automation, but time of computation still very long, reason is due to if/want human program after, again/then delivered into machine to operate. These data and program after passing through process of human, again difficult to avoid not give rise to error. Present time this kind error still is very large problem, therefore for the sake of cause process of computation reliable also high speed, machine automatically program is very important. This is one subject which present time programmers concentratedly study.

(3) research and development of part and component. If cause computer satisfy above mentioned requirement, aside from

make effort from logic structure and program aspect, still if from machine itself and its building block, then possible solve problem. Therefore research of part and new component is very important. Computation technique if/want used for two kind radio component, that is active elements and passive elements. Active elements, speaking from work speed, dimension, weight, consumption efficiency, life and reliability and so on, electron tube all not compare favorably with transistor. Presently already has transistor circuit

of computer after work 26,000 hour not yet discover deterioration phenomena. Compared with life of electron tube several hundred, several thousand hour, reliability of transistor circuit is greatly has elevated. Therefore transistor is determine if/want direction of development. Transistor at frequency, efficiency property and temperature special characteristic and so on aspect, present time still not possible completely satisfy requirement, require continuously study and improve. Passive elements include electric resistance, capacitance, pulse transformer, delay lines, storage component, logic component, contacts and so on. From the viewpoint of reliability, requires component in possible utilize condition (high temperature, high pressure, strong vibration, radiation and so on) and time operation reliable, value stable. From the viewpoint of efficiency and high speed, miniaturization of component is very important, but especially important is research of new type component. Use solid-state component circuit which ferromagnetic material, ferroelectric material, super-conductor material and so on accomplish is new development direction. Present time use ferromagnetic material to make memory component possesses undebatable merits. Almost all computer which recent manufactured, use magnetic core component to make high speed storage device. Must suppose

switching time limit of magnetic core component in the neighborhood of microsecond, thus, use access period which it has made then limit depend upon/at several microsecond. Recently discover switching time of magnetic core possible shorten, therefore possible use it in order to make storage device of even higher speed. Of course, multi-path magnetic core, magnetic film, magnetic rod and so on new component still is one principally research direction. Use magnetic core to make logic component, especially variable parameter component is that which worth study.

Ferroelectric component is use electric field in order to that which operate, therefore it is one kind electric potential type component. On this point that of vacuum tube same type and it is. And that is to say they used together comparatively suitable. Ferrromagnetic similar to super-conductor component is use electric current product magnetic field in order to operate, therefore they is electric current type component, this is that of similar to transistor . Thus, use high efficiency transistor in order to directly drive magnetic core is comparatively good match. Merits of super-conductor component is due to it dimension very small and consumption efficiency small. But ferroelectrics and super-conductor component still subject to definite limit, not possible with great quantity apply Secondary Prefix Error Final Pass If can find one kind property (shape of hysteresis loop, temperature, frequency and so on) good but manufacture easy ferroelectric material, or under ambient temperature (or not low temperature) has material of super-conducting phenomena, they all possible rapidly on computation circuit was utilized. Research of present time they work still mostly at property of material direction.

**APPENDIX III**  
**Project Personnel**

## PROJECT PERSONNEL

The first phase of the work was a cooperative effort by personnel from IBM Research, the Mergenthaler Linotype Company and staff members of the Yale University Institute for Far Eastern Studies. The second phase of the Contract work was performed entirely by IBM Research Staff Members with Mr. Fang-Yu Wang acting as linguistic consultant.

### IBM Research Personnel

#### Experimental Systems Research Department

H. W. Chang  
S. S. Soo  
R. D. Villani  
A. V. Michaelson  
M. K. Schneider

### Mergenthaler Linotype Company Personnel

Merritt Brown  
Edward Bechtold (Optical Consultant)  
Franz Fischer  
James Monagle  
Andrew K. Wong

### Yale University Institute for Far Eastern Studies Staff Members

Henry Fenn  
Gerald Kok  
Robert Tharp  
Fang-Yu Wang  
James Wrenn