

# 汉字精密照相排版系统

张炘中 夏莹 何克忠<sup>1)</sup>

(清华大学)

## 摘 要

本文提出了计算机控制的精密照相排版系统。该系统包括输入汉字大键盘、自动照相排版机及由用线性 E 型磁芯组成的  $32 \times 32$  点阵汉字发生器和激光扫描点阵式汉字印字机组成的校对系统。

为保证文字和版面质量、较高的照排速度和充分利用汉字使用频度差别大的特点,自动照相排版机的工作原理是用一个大透镜与几百个小透镜相配合,将在字模盘上不同位置的汉字成象在一点上。该机有宋、仿宋、楷和黑四种字体,具有十级变倍,七种变形,相当于二百万铅字。

## 一、前 言

随着计算技术的发展,汉字信息的获取、传递、存贮、变换、通讯以及输出信息的处理技术正在蓬勃发展<sup>[1,2]</sup>。国内在汉字照相排版、汉字邮电通讯、汉字终端、汉字输入编码等方面开展了研究工作,各种输入键盘、汉字发生器、印字机、显示器、照排机及应用软件相继出现,为我国汉字信息处理技术的进一步发展打下了基础。

本文提出了一个用计算机控制的包括汉字输入、校对、编辑、组版及输出的汉字精密照排系统。

## 二、汉字的特点和照排质量要求

设计一个照排系统必须考虑汉字的特点及印刷的质量要求。汉字的特点有:

1) 汉字是方块字,字量多,一般认为有 5—6 万字。当前的通用字也有 6 千多。汉字有宋、仿宋、楷黑等字体,每种体又有多种字号,因此汉字库的容量十分庞大。

2) 各个汉字出现的概率(使用频度)差别大。一九七六年底新华印刷厂等单位就当时流通的书籍、报刊,统计了 21,629,372 个字的使用频度。表 1 为不同汉字集出现概率分布。其中,  $n$  为以概率大小排的序号,  $P_n$  为  $n$  号汉字集出现的概率。可见一个最常用字“的”出现的概率  $P_1$  就占 4%。如果把“的”与不常用字集出现的概率相比较,有

$$P_1 = 400 \times (P_{6374} - P_{5200}),$$

本文于 1981 年 6 月在中国中文信息研究会第一次学术交流会上宣读,修改稿于 1981 年 7 月 10 日收到。

1) 参加本系统工作的有: 清华大学、北京新华印刷厂、长春光机所、邮电 522 厂、南京模具厂、北京新华字模厂等单位。

表 1 不同汉字集出现概率分布表

$n$	1	162	942	2399	2958	3838	5200	6374
$P_n$	0.04	0.50	0.90	0.99	0.996	0.999	0.9999	1.0000

即“的”一个字的使用频度比概率最低的 974 字集的整体使用频度还高 400 倍!

出现概率高的字往往是构词能力强的字。照排用字量、输入键盘的布置、盘外字的选择等都要充分利用汉字出现概率差别大的特点进行设计,以便提高照排速度。

3) 汉字字形复杂,有笔锋,因此要有足够的解象力才能达到印刷制版的要求。

汉字照排质量包括单个文字质量和整个版面的质量。表 2 和表 3 列出了用于不同场合的汉字输出质量要求,供参考。

表 2 单 字 质 量

项 目 \ 使用类别	监 视	简易印刷 (印字机)	出版印刷 (照排机)
解象力(线/毫米)	4—8	8—16	$\geq 20$
密 度	2.0—2.5	2.2—2.5	$2.5 \pm 0.1$
保 真 度	低	中	高

表 3 整体版面质量

项 目 \ 使用类别	监 视	简易印刷 (印字机)	出版印刷 (照排机)
文字定位精度 (对 $4 \times 4 \text{mm}^2$ 字)	$\leq 1.0$	$\leq \pm 0.2$	$\pm 0.03 - \pm 0.05$
一行字高低允差 (mm)			
左右允差 (mm)		$\leq \pm 0.3$	$\leq \pm 0.08$
版面密度均匀性	$\leq 0.6$	$\leq 0.4$	$\leq 0.3$

### 三、照排系统概貌

#### 1. 系统设计的三个要点

1) 精密汉字发生器方案的选择是设计精密照排系统的一个困难任务。若采用字模数字化方案,当采用四种字体、十种字号、六千个字量,则一整套字模就是  $2.4 \times 10^5$  个。要达到 25 线/毫米的解象力,约要  $4.5 \times 10^9$  位存储容量,即使采用高倍信息压缩技术(以不降低质量为前提),存储容量也很庞大,这就增加了系统的成本。为了既能产生  $2.4 \times 10^5$  个合格字模,又能降低造价,作者采用字模盘存储汉字,用光学方法改变字号的光机式照排机方案,能在较快的照排速度下输出合格的版面。

2) 为提高整个照排系统的效率,研制了数字存储汉字发生器,并设计了校对修改分系统。该分系统采用汉字输出速度比主系统高一个量级的激光印字机,在普通纸上印出汉字,修改由编辑软件完成。

3) 汉字输入一直是汉字信息处理的一个关键问题。在汉字印字机、显示器、照排机等设备已趋成熟的今天,这个问题尤其突出。以手动输入为主的输入方案五花八门,种类繁多。不同的使用领域要选择不同的输入方案。例如在计算机汉字终端中,应以使用方便、记忆量少的小键盘为主。为在错码率相同的前提下提高输入速度,本系统采用有区别键的汉字大键盘。

### 2. 系统组成

精密照排系统由汉字键盘穿孔机、汉字印字机、汉字发生器、控制计算机、精密输出自动照排机(包括照排控制机)及组版、编辑等软件组成。见图 1。

工作流程为,将批注后的原稿用汉字键盘穿孔机穿孔成字符代码纸带,然后输入汉字印字机印出校样,经人工校对后再用汉字键盘穿孔机穿出修改纸带,将修改纸带与原稿纸带一起输入计算机进行编辑、组版等处理,再由照排控制机控制自动照排机拍照得到组版胶片,最后拼版制成感光尼龙版上机印刷。

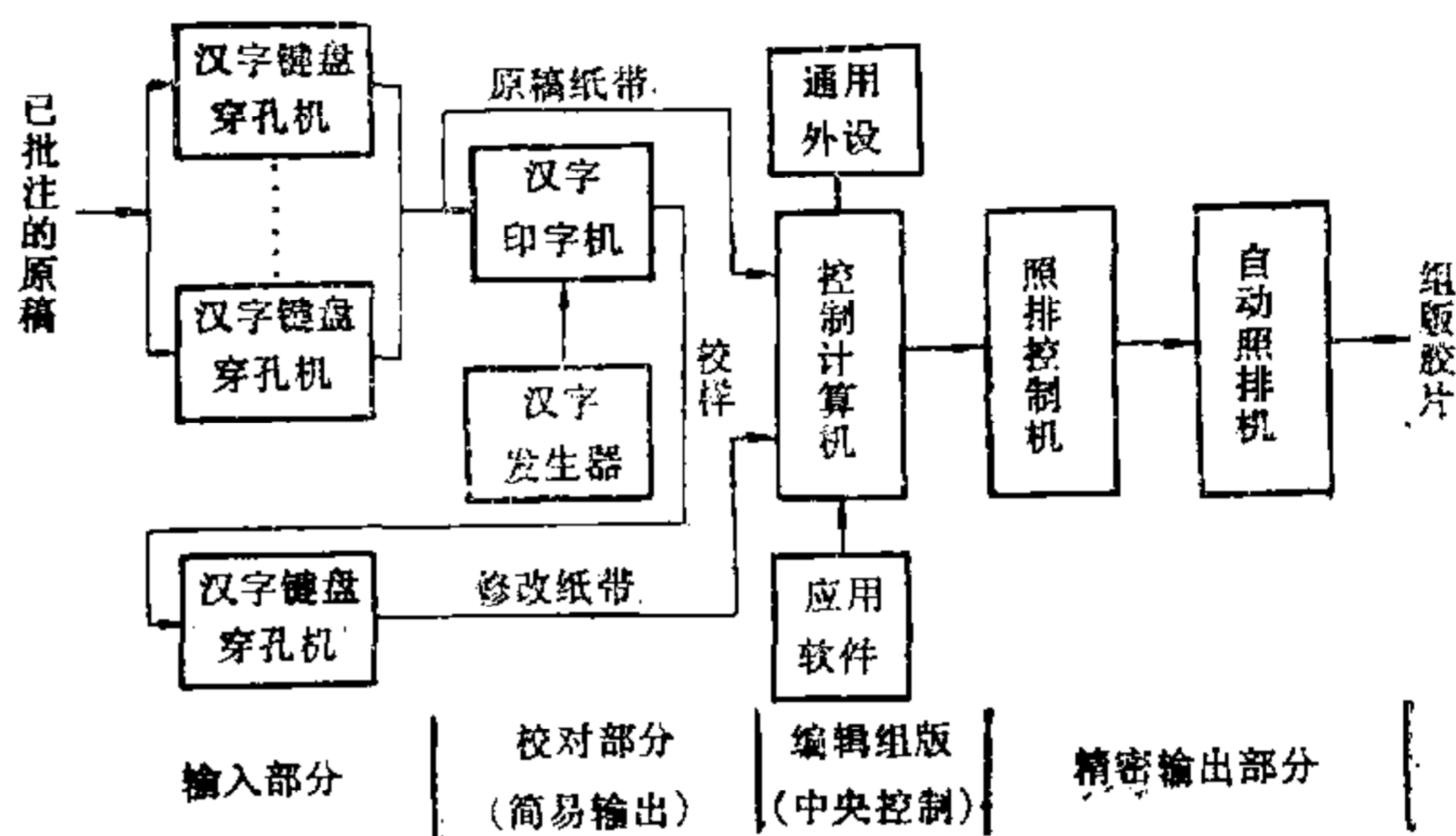


图 1 汉字精密照排系统方块图

### 3. 照排系统各部分概述

除自动照排机(包括照排控制机)专门叙述外,系统其余部分介绍如下。

1) 输入键盘用整字键盘,有汉字键六百多个,每键有九个字。除此以外,还有九个区别键置于键盘左侧。盘面上还设置了约一百个功能键。

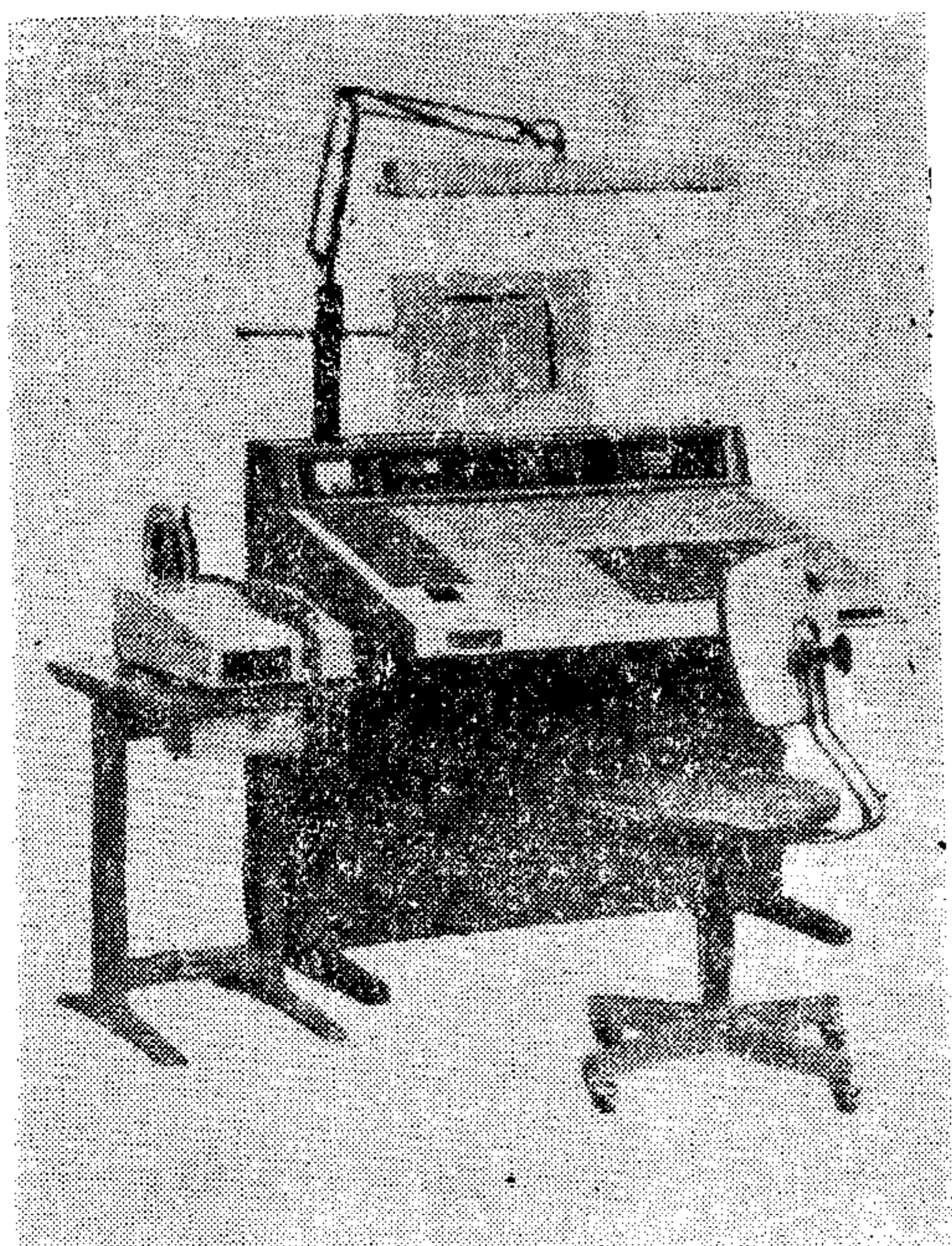


图 2 整字键盘穿孔机

盘面按汉字出现概率和词组联接分区,把概率大的一千六百多个字分布在盘面下部中右区(图 2 中深色部分)以便操作。其他文字按偏旁部首排列分区,以利寻找。由于这一千六百多个字出现的概率已占百分之九十

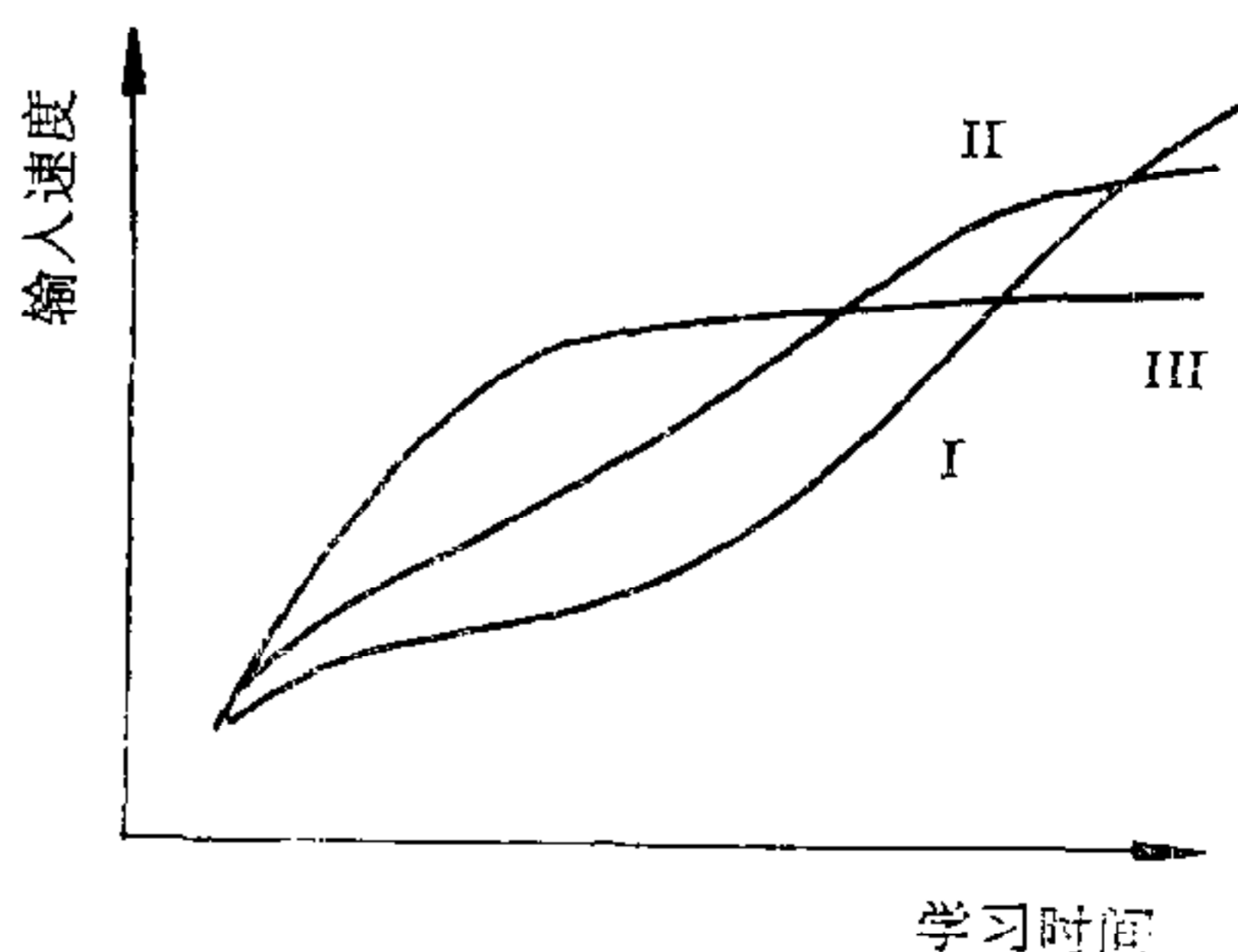


图 3 不同类型键盘的学习曲线

六以上,所以,盘面虽大却很少影响输入的平均速度,一般来说熟练的操作员每分钟输入字数可达 60 个以上。

这种带区别键的大键盘与字表触笔式大键盘或小键盘相比,对熟练的操作员来说,可以在长时间内保持较高的输入速度,且不易感到疲劳.

不同类型键盘的学习曲线见图3.(图中I为本键盘学习曲线.)

汉字输入代码为七单位双字节.

2) 汉字印字机和发生器. 由于汉字字量很大,因此,汉字印字机几乎都采用点阵式. 为了提高输出速度,我们选择了激光扫描印字方案. 它与静电、喷墨、电子照相等高速印字机相比,从发展前途来说,有以下突出优点: ① 文字质量高; ② 印字速度快; ③ 技术较简单; ④ 利用激光的强功率,存在直接制版的可能性. 缺点是感光鼓的寿命不长,要经常维修.

图4为包括汉字发生器在内的激光扫描汉字印字机构成图. 该机对 $4 \times 4\text{mm}^2$ 的文字,在普通纸上可以 $\geq 200$ 字/秒的速度印出分辨率为8线/毫米的文字,可模拟实现另段、另页,另面的功能,能自动印出行、页标.

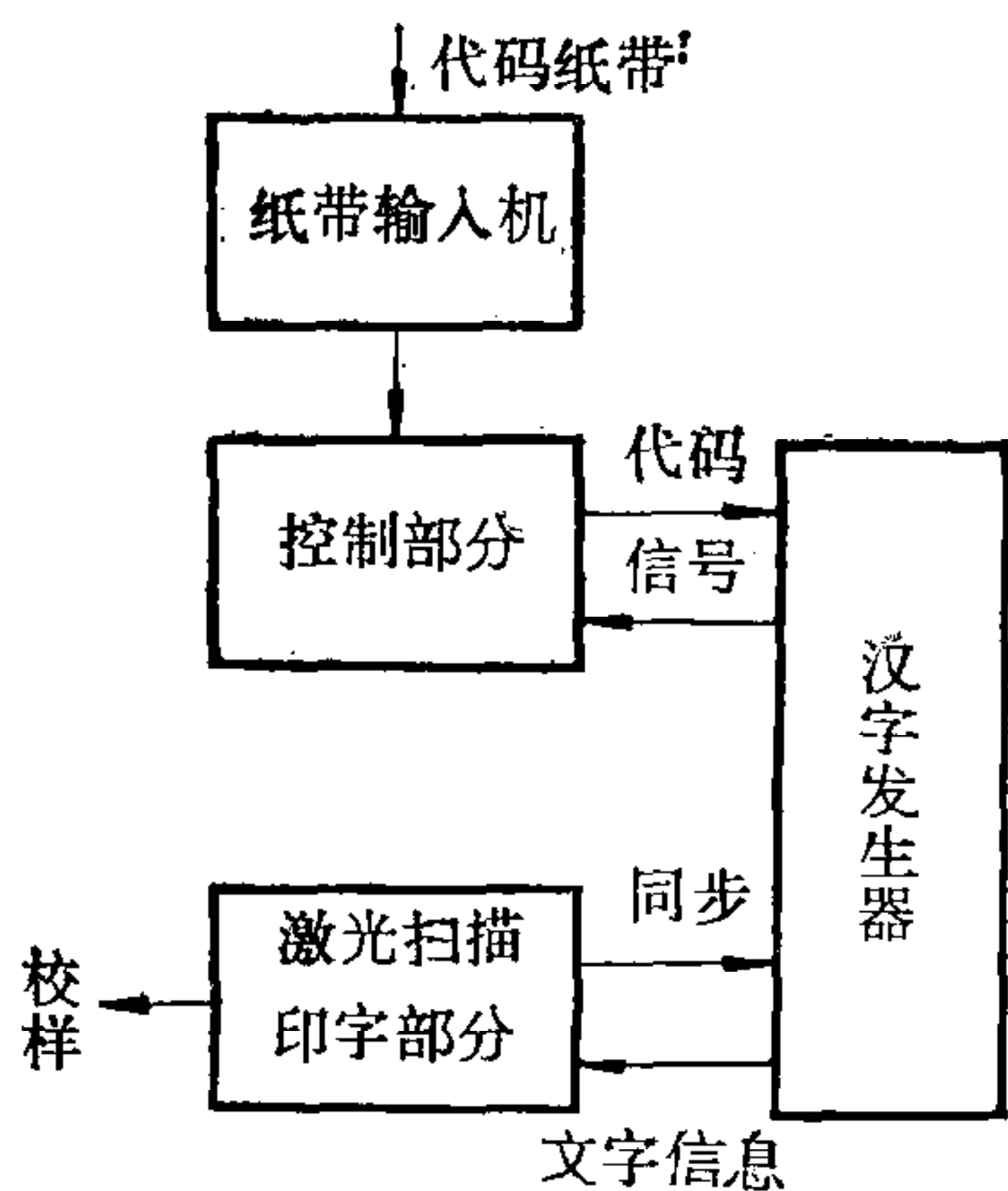


图4 激光扫描汉字印字机的构成

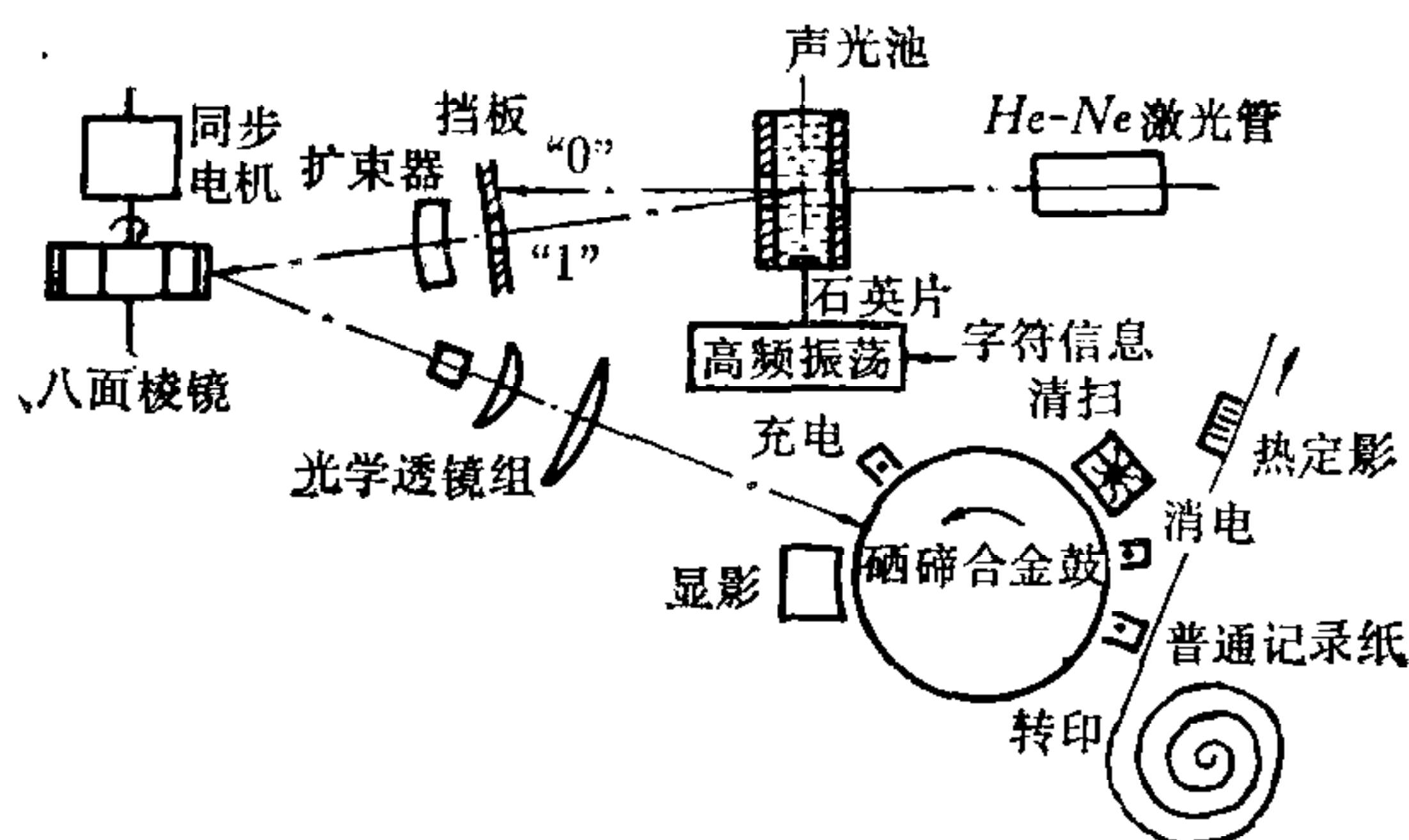


图5 激光扫描、印字部分原理示意图

图5为印字机扫描、印字部分原理图. 光源为氦-氖激光,八面棱镜及硒鼓的同步转动分别完成横向和纵向扫描. 字符信息加到高频振荡器上,通过声光池来调制激光束. 预先充上静电的硒鼓在激光扫描后形成静电潜像,显影后转印到普通纸上.

汉字发生器<sup>[3]</sup>为印字机提供文字信息. 汉字数字存贮的原理如下:

设汉字图形为矩阵

$$P = (p(i, j)), i, j = 1, 2, \dots, N, \text{ 取 } N = 32.$$

其中,

$$p(i, j) = \begin{cases} 1 & (\text{文字上}) \\ 0 & (\text{空白}). \end{cases}$$

设汉字库有 $Q$ 个字,则有 $P_l = (P_l(i, j)), l = 1, 2, \dots, Q, \text{ 取 } Q = 6161.$  用E型线性变压器(磁蕊)存储数字化后的汉字信息. 由于有效地排除了干扰,使信息错误率远小于0.01%. 汉字发生器已经过鉴定,其主要技术指标是: 储字符量为6161个;文字质量为 $32 \times 32$ 点阵;文字发生速度 $\geq 4\mu\text{s}/\text{字线}$ ;字体为宋体;输出方式为按字线(32位)串行或并行输出.

图 6 为示波器屏幕上显示的汉字发生器发生的部分字样。

3) 控制计算机采用 DJS-130 机, 内存为 32K, 存放应用软件和待照文字代码, 通过专用接口用程序中断方式控制照排机。

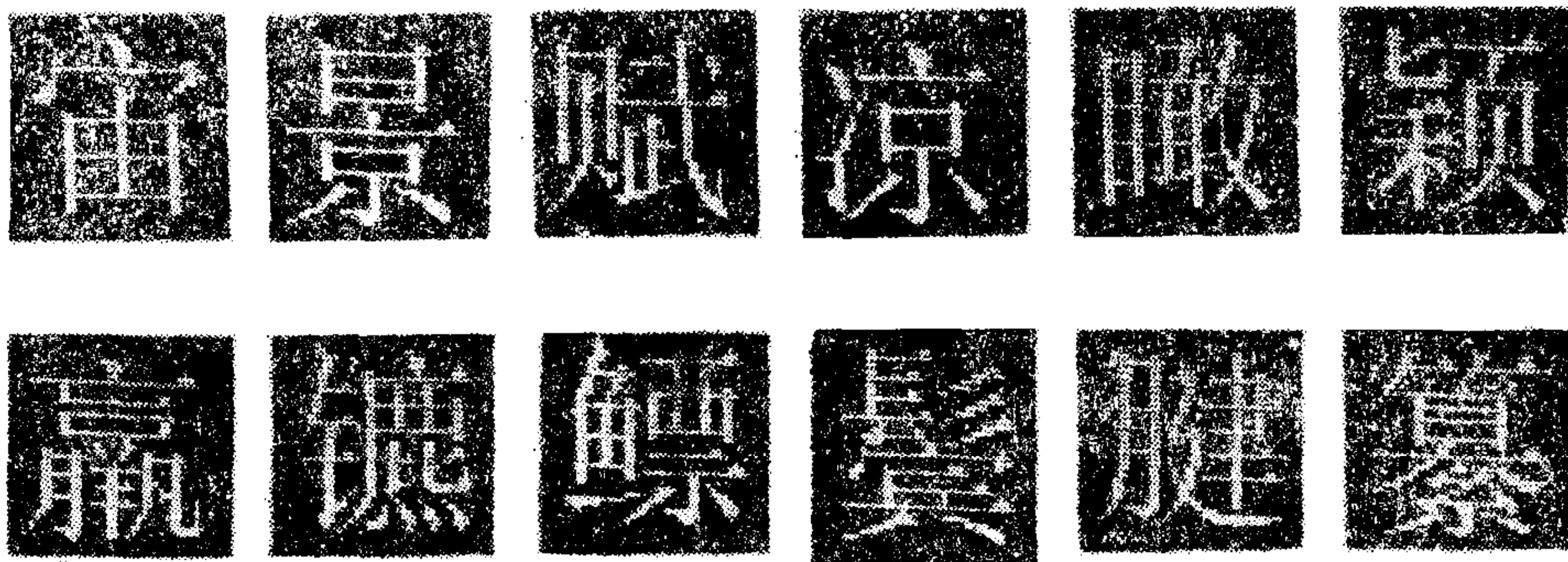


图 6 示波器屏幕上显示的汉字发生器发生的字样

#### 四、自动照排机原理

国外光机式照排机大多采用旋转字模筒(或盘), 用闪光灯动态拍照, 日本富士通的 F6531C 是个典型例子<sup>[4]</sup>。其缺点是不能充分利用汉字出现概率差别大的特点提高照排速度。本文提出一种用平面字模板, 由小透镜合像, 用连续光源静态拍照的方案。

##### 1. 选字原理

用矩阵  $A$  表示字模板上字模的分布,  $a_{ij}$  为字模元素, 有  $A = (a_{ij})$ ,  $1 \leq i, j \leq n$ 。为缩短字模板运动的距离以提高照排速度, 把  $A$  矩阵分成若干分块矩阵,  $A = (A_{\alpha\beta})$ ,  $1 \leq \alpha \leq s$ ,  $1 \leq \beta \leq t$ 。子块  $A_{\alpha\beta}$  为  $n_\alpha \times n_\beta$  矩阵,  $n_\alpha = \frac{n}{s}$ ,  $n_\beta = \frac{n}{t}$ 。即  $A_{\alpha\beta} = (a_{gh})$ ,  $1 \leq g \leq n_\alpha$ ,  $1 \leq h \leq n_\beta$ 。  $i, j$  和  $\alpha, \beta$  的关系为  $(\alpha - 1) \frac{n}{s} + 1 \leq i \leq \alpha \frac{n}{s}$ ;  $(\beta - 1) \frac{n}{t} + 1 \leq j \leq \beta \frac{n}{t}$ 。若取  $n = 60$ ,  $s = t = 15$ , 则  $n_\alpha = n_\beta = 4$ 。一块字模板上可容纳  $\leq n \times n = 3600$  个字模。用 10 块字模板可容纳  $3.6 \times 10^4$  个字模。

设置  $s \times t$  个小透镜, 每个小透镜上加一个与字模子块  $A_{\alpha\beta}$  相对应的光栏, 控制光栏的开合可从  $s \times t$  个字模子块中任选一个字模子块  $A_{\alpha\beta}$ , 然后移动字模板, 从子块  $A_{\alpha\beta}$  里的  $n_\alpha \times n_\beta$  个字模元素中选出需要的字模  $a_{gh}$ 。另外, 把出现概率大的字放在不要移动字模板的地方, 而把出现概率小的字放在要移动字模板最远方能找到的地方, 从而大大提高了选字速度和平均照排速度。

本文作者还设计了多块字模板的自动换板系统, 实现了用多块字模板存贮字模。

##### 2. 合像原理及照排原理的改进

连续光源经聚光后变为平行光均匀照射在字模板上, 经小、大透镜系统把  $s \times t$  个字成像在同一焦面  $N - N$  处,  $N - N$  为第一像面。控制字模板的移动使被选字总是处在

小透镜的中心，经过变倍光学部件，使置于它场镜处的第一像面的像在胶片上第二次成

像(见图7)。变倍光学部件可改变字号10级以上。

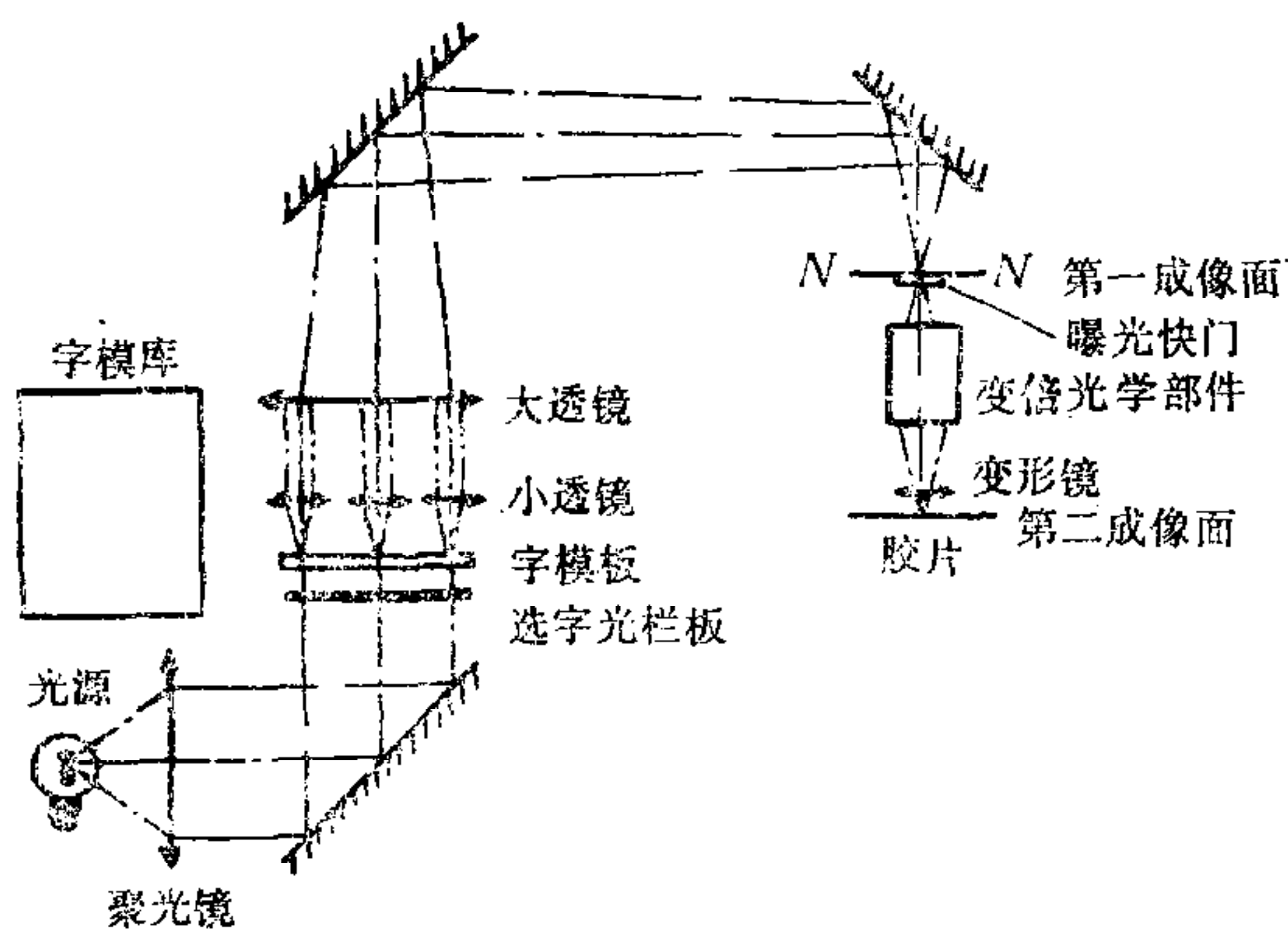


图7 自动照排机成像原理图

为充分利用不同汉字出现概率差别大的特点，进一步提高照排的平均速度，可把单字模板扩展为静、动两个字模板，用双光路、双暗盒交替拍摄。除了小、大透镜合像外，还用半反半透镜把两个字模板的像合在一起，同时又形成后半部的双光路。静止的字模板上放置出现概率最高的 $s \times t$ 个字，可动的字模板上放置 $s \times t \times n_a \times n_b$ 个字，则共有 $2s \times t$

个字不需移动字模板，如 $s \times t$ 取553, 1106字集出现的概率就为0.92，这意味着选字时字模板几乎不动，因而进一步提高了照排速度。

### 3. 照排控制机

照排控制机由接口和控制系统两部分组成。

1) 接口的作用是接收计算机发出的一个或多个十六位信息(高四位为特征位，低十二位为次数、长度、方向等的控制码)，经判别处理后分配给自动照排机的某系统，并把自动照排机的回答传给计算机。自动照排机每项动作完成后均向计算机请求中断。

2) 控制系统包括：用来控制字模板和暗盒移动的数字快速位置控制系统(闭环)<sup>[5]</sup>；用来控制字模库、移行、变倍、变形等运动的步进电机控制系统；用来控制曝光快门的快速步进电机控制系统<sup>[6]</sup>以及用来控制选字光栏开合的快速电磁铁系统。

外面，在有些模糊的玻璃窗外面，一对对男女手挽着手地走过，他们向四面八方走去，有的由地下铁道走出来，有的是走进去，有的从公共汽车上下来。

“异国的生活，”他缓慢地说道。

沉默了许久，没有人说话。

“无论如何，我从来没有丧失信心，”他继续说道，“我相信共产主义的内在力量。如果他们对我说，‘这就是你的生活，你过去的生活的篇章，一个假名字就是一个官衔，你为自己选择一部分，把其余的留给阶级的事业！’我为自己什么也不要，或者更确切地说，我只取出其中的几天，那就足够了。”

“你一生中有没有为自己花去些时日？”我问他道。

“没有，除了少有的几个星期日，”他说道。

图8 自动照排机样机照排的部分字样

## 五、结 论

本文提出了一种独特的自动照排机原理,介绍了包括汉字印字机在内的一个精密照排系统,经多年试生产实践,已照排了书刊、画报、辞典、表格等共约 300 余万字,质量基本符合出版要求(参看图 8)。实践证明本文所述自动照排机方案是可行的。据改进原理设计成的照排机正在进行调试。

## 参 考 文 献

- [1] 松田功等,漢字処理とハ——ドウエア, *FUJITSU*, 27 (1976), 447—458.
- [2] 米村一介,サンケイ新聞社の漢字処理システム, *FUJITSU*, 27 (1976), 416—427.
- [3] 张忻中等, 32×32 点阵汉字发生器的试验研究, *清华大学学报*, 20(1980), 45—59.
- [4] 塚谷等, FACOM 6531C 全自动写真植字装置, *FUJITSU*, 21 (1970), 99.
- [5] 刘植楨等, 提高伺服系统快速性的探讨, *清华大学学报*, 20(1980), 99—109.
- [6] 夏莹等, 反应式步进电机的快速控制问题, *自动化学报*, 6(1980), 302—310.
- [7] 海住嘉彦等, FACOM 6533A 写真植字装置, *TUJITSU*, 30 (1979), 173—184.

# A PRECISE CHINESE CHARACTER PHOTOTYPESETTING SYSTEM

ZHANG XINZHONG XIA YING HE KEZHONG

(*Qinghua University*)

## ABSTRACT

A precise chinese character phototypesetting system controlled by computer is presented. It consists of a big keyboard of chinese characters, automatic phototypesetting machine and a check system consisting of a chinese character generator and a laser scanning dot-matrix printer. The chinese character generator is composed of 32×32 dot-matrix implemented with E-shape soft magnetic-core.

In order to guarantee the quality of characters and page format, to have a higher speed of phototypesetting and to make full use of large difference in the frequency of usage of chinese characters, the working principle of this automatic phototypesetting machine is to focus the image of all chinese typeheads at different position on the model plate into a small square by means of one big lens and several hundred small lenses.

This machine has four typefaces: Song-dynasty-style typeface, imitation song-dynasty-style typeface, blackface type and regular script. There are 10 sizes and 7 shapes for each typeface, so that the machine may be considered as consisting of about 2 million typecasting.