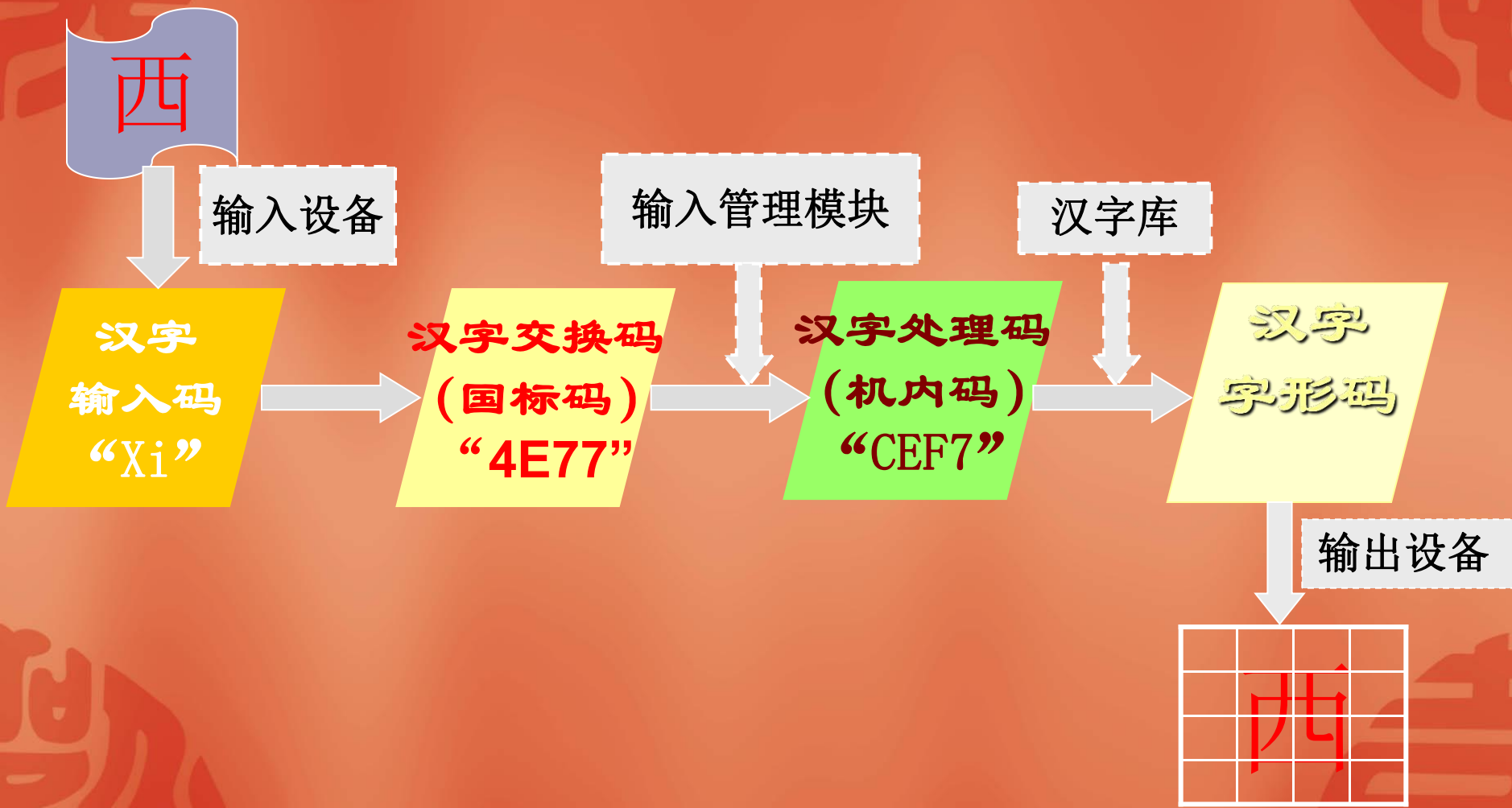


2.3.3 非数值信息编码_汉字编码

- ◆ 汉字的特点
 - ✦ 图形文字；
 - ✦ 常用汉字多；
 - ✦ 形状和笔画差异很大；
- ◆ 决定了汉字字符的编码方案必须完全不同于西文的编码方案；
- ◆ 必须解决汉字的输入编码、存储编码、显示和打印字符的编码问题。

2.3.3 汉字编码_汉字处理转换示意图



2.3.3 汉字编码_汉字输入码

◆ (1) 输入码：外码

- ★ 输入码是解决汉字输入采用的编码，如“全拼”、“双拼”、“五笔”、“智能ABC”等输入编码。
- ★ 虽然每一种输入法对统一汉字的编码各不相同，但经过转换后，存放在计算机内的编码都是汉字机内码。
- ★ 哪种输入码最好，回答是“你熟练掌握的那种输入码就是最好的”。



音码：按照汉字字音特征编码（如智能ABC、拼音输入法等）

形码：按照汉字的形体结构特征编码（如五笔字型等）

音形码：以字音为主，辅以字形特征的编码

字形码：以字形为主，辅以字音特征的编码

2.3.3 汉字编码_交换码

◆ (2) 交换码:

- ★我国在1981年制定了《信息交换用汉字编码字符集·基本集》作为国家标准GB2312-80编码字符集，称为区位码
- ★每个汉字采用两个字节表示。分别表示其所在的区号（两位十进制）和位号（两位十进制）。
- ★GB2312-80规定，所有的国标汉字和符号组成一个94*94的矩阵，该矩阵中的每一行称为一个“区”，每一列称为一个“位”。

2.3.3 汉字编码_国际码

汉字的**区位码**：区号（行） + 位号（列）

国际码：区号和位号加上十进制数32后得到的结果对应的二进制代码为其国际码。

即：**国际码=区位码+2020H**

例如，“啊”的区位码：16、01。

二进制：

0	0	1	1	0	0	0	0	0	0	1	0	0	0	0	0	1
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---

十六进制：

3	0	2	1
---	---	---	---

“啊”的国标码： $(3021)_{16}$

2.3.3 汉字编码_处理码

- ◆ (3) 处理码 (机内码):
- ◆ 计算机系统内部存储、处理和传输汉字所使用的代码。
- ◆ 机内码: 常用十六进制编辑工具有 **UltraEdit(UE)**和**WinHex**, 用来查看文字内码
 - ✦ 将汉字区位码前后两个字节的最高位置**1**, 即为机内码。

例如, “啊” 的国标码: $(3021)_{16}$

二进制:

0	0	1	1	0	0	0	0	0	0	1	0	0	0	0	0	1
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---

机内码:

1	0	1	1	0	0	0	0	1	0	1	0	0	0	0	0	1
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---

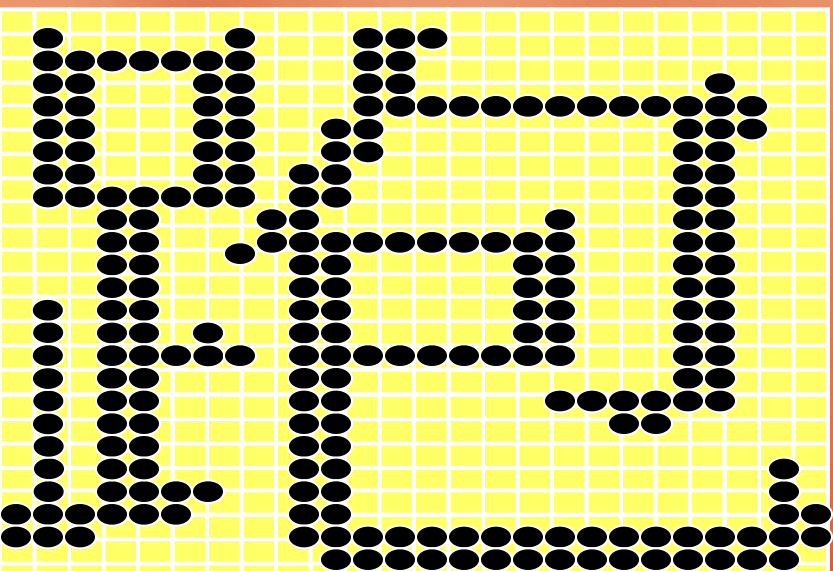
“啊” 的机内码: $(B0A1)_{16}$

2.3.3 汉字编码_汉字字形码

(4) 汉字字形码：为了汉子的输出显示和打印，需要描述汉字的字形，汉子的字形通常有两种表示方式：点阵方式和矢量方式。

汉字显示的时候是以字形码显示的，字形码有两种，点阵和矢量。点阵就是有笔画的地方有黑点，没笔画的地方有白点，原理类似于位图，因此放大后，会有小方块出现。而矢量方式显示的汉字则不会变模糊，还是那么圆滑，类似于矢量图。

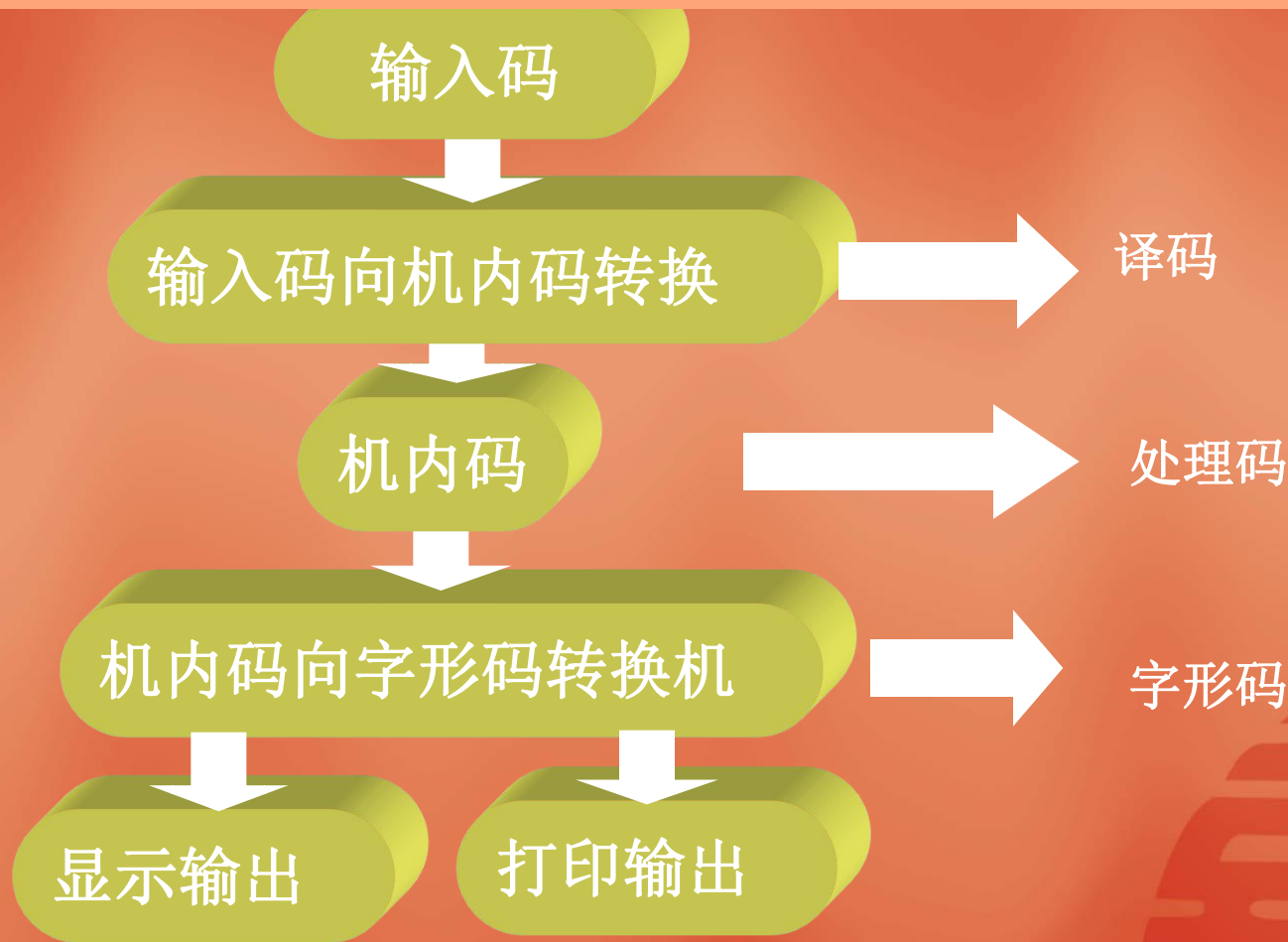
那么，想要看到二者的不同，其实很简单，在C:\WINDOWS\Fonts，里面有多种字体，其中扩展名为TTF的是矢量字库，扩展名为FON的，是点阵字库，可以看到，基本上都是矢量字库



汉字点阵类型	点阵	占用字节数
简易型	16×16	32
普及型	24×24	72
提高型	32×32	128
精密型	48×48	288

2.3.3 汉字编码_汉字处理转换示意图

➤ 几种汉字码的对应关系是:



练习

1、五笔输入法按输入的相关特征分类，属于（**D**）

A 音码 B 形码 C 音形码 D 形音码

2、大写A的ASCII码值是65（十进制），则字母D在UE软件中显示的内码是（**B**）

A 43H B 44H C 45H D 67H

3、汉字“人”的内码是“11001000 11001011”，那么在UE软件中观察到的十六进制编码是（**B**）

A B8 CB B B8 BA C D8 DC D C8 CB

4、汉字的点阵码是一种用黑白两色点阵来表示汉字字形的编码。一个16*16的点阵字模的存储容量是（**C**）字节

A 1 B 16 C 32 D 64