

# 第2章 数据的表示和运算



中华人民共和国成立70周年



1949 — 2019  
大连理工大学 建校70周年

主要内容:

## (一) 数制与编码

1. 进位计数制及其相互转换
2. 真值和机器数
3. BCD 码
4. 字符与字符串
5. 校验码

## (二) 定点数的表示和运算

1. 定点数的表示：无符号数的表示；有符号数的表示。
2. 定点数的运算：定点数的位移运算；原码定点数的加/减运算；补码定点数的加/减运算；定点数的乘/除运算；溢出概念和判别方法。

## (三) 浮点数的表示和运算

1. 浮点数的表示：浮点数的表示范围；IEEE754 标准
2. 浮点数的加/减运算

## (四) 算术逻辑单元 ALU

1. 串行加法器和并行加法器
2. 算术逻辑单元 ALU 的功能和机构

海纳百川 自强不息 厚德笃学 知行合一



# 2.1 数字信息和二进制编码



中华人民共和国成立70周年



1949 — 2019  
大连理工大学 建校70周年

**信息：数值、文字、符号、语音和图形、图像等。**

➤在计算机内部，信息都必须用数字化的形式被存储、加工和传送，不同信息要通过编码来表示

➤ **计算机中普遍选用两个基本符号： 1 0**

**优点：** 基本符号个数最少，物理上容易实现；二进制码表示数值数据运算规则简单；与二值逻辑的真、假两个值对应简单

**编码：用少量简单的基本符号，对大量复杂多样信息进行一定规律的组合表示。**

➤ 编码的两个要素：少量简单的基本符号、一定的组合规则

➤ 数值数据编码需解决：

- ✓ 数的大小：采用什么进位计数制
- ✓ 数的符号表示：**原码、补码、反码**
- ✓ 小数点的表示：定点、浮点

11合一



## 2.1.1 进位计数制及其相互转换



中华人民共和国成立70周年



1949 — 2019  
大连理工大学 建校70周年

### 1. 进位计数制

- 是指按照进位制的方法表示数，不同的数制均涉及两个基本概念：基数和权。
  - ✓ 基数：进位计数制中所拥有数字的个数。
  - ✓ 权：每位数字的值等于数字乘以所在位数的相关常数，这个常数就是权。
- 任意一个r进制数N，设整数部分为n位，小数部分为m位，则N代表的实际值可表示为：

$$N = D_{n-1}r^{n-1} + D_{n-2}r^{n-2} + \dots + D_0r^0 + D_{-1}r^{-1} + D_{-2}r^{-2} + \dots + D_{-m}r^{-m}$$
$$(N)_r = \sum_{i=-m}^{n-1} D_i r^i$$

附赠“日积月累”与“合二为一”



# 进位计数制



中华人民共和国成立70周年



1949 — 2019  
大连理工大学 建校70周年

## 计算机中常用的 4 种进位数制 ( $r$ —数制的基)

二进制(B):  $r = 2$ , 基本符号: 0 1

八进制(O):  $r = 8$ , 基本符号: 0 1 2 3 4 5 6 7

十进制(D):  $r = 10$ , 基本符号: 0 1 2 3 4 5 6 7 8 9

十六进制(H):  $r = 16$ , 基本符号:

0 1 2 3 4 5 6 7 8 9 A B C D E F

其中 A~F 表示十进制数 10~15

➤ 4 种进位数制之间的关系:

- ✓ 二进制用于计算机内部;
- ✓ 八和十六进制是二进制的缩写;
- ✓ 十进制用于编程、操作人员。

防泄密 | 自主研发 | 产学研用 | 深度融合



## 2. 不同数制间的数据转换

### ➤ 二、八、十六进制数转换成十进制数

利用上面的公式进行计算：

$$(N)_2 = \sum D_i \cdot 2^i, (N)_8 = \sum D_i \cdot 8^i, (N)_{16} = \sum D_i \cdot 16^i$$

### ➤ 十进制数转换成二进制数

通常要对一个数的整数部分和小数部分分别进行处理，各自得出结果后再合并。

■ 对整数部分，**一般采用除2取余数法**，其规则如下：

- 将十进制数除以2，所得余数（0或1）即为对应二进制数最低位的值。然后对上次所得商除以2，所得余数即为二进制数次低位的值，如此进行下去，直到商等于0为止，最后得的余数是所求二进制数最高位的值。

■ 对小数部分，**一般用乘2取整法**，其规则如下：

- 将十进制数乘以2，所得乘积的整数部分即为对应二进制小数最高位的值，然后对所余数的小数部分部分乘以2，所得乘积的整数部分为次高位的值，如此进行下去，直到乘积的小数部分为0，或结果已满足所需精度要求为止。

# 思考题:



➤ 请分别将  $(125)_{10}$  和  $(0.8125)_{10}$  化成二进制。

海纳百川 自强不息 厚德笃学 知行合一



# 十进制整数转换成二进制



中华人民共和国成立70周年



1949 — 2019  
大连理工大学 建校70周年

十进制转换为二进制方法：

$$125 = 1 \times 2^6 + 1 \times 2^5 + 1 \times 2^4 + 1 \times 2^3 + 1 \times 2^2$$

$$+ 0 \times 2^1 + 1 \times 2^0$$

$$125 = 111\ 1101\text{B}$$

	余数
2   125	1
2   62	0
2   31	1
2   15	1
2   7	1
2   3	1
	1

低位

高位

海纳百川 自强不息 厚德笃学 知行合一





# 十进制小数转换成二进制



中华人民共和国成立70周年



1949 — 2019  
大连理工大学 建校70周年

0.8125的二进制

$$0.8125 \times 2 = 1.625 \quad \text{取整是} 1$$

$$0.625 \times 2 = 1.25 \quad \text{取整是} 1$$

$$0.25 \times 2 = 0.5 \quad \text{取整是} 0$$

$$0.5 \times 2 = 1.0 \quad \text{取整是} 1$$

高位

低位

即0.8125的二进制是0.1101 (第一次所得为最高位,最后一次得到为最低位)

海纳百川 自强不息 厚德笃学 知行合一







## ➤ 二进制数、八进制数和十六进制数之间的转换

八进制数和十六进制数是从二进制数演变而来的：  
由**3**位二进制数组成**1**位八进制数；  
由**4**位二进制数组成**1**位十六进制数。

- 对于一个兼有整数和小数部分的数以小数点为界，小数点前后的数分别分组进行处理，不足的位数用**0**补足。
- 对整数部分将**0**补在数的左侧，对小数部分将**0**补在数的右侧。这样数值不会发生差错。

海纳百川 自强不息 厚德笃学 知行合一



# 二进制数与八、十六进制数之间的直接转换法

中华人民共和国成立70周年 1949—2019 大连理工大学 建校70周年

二<—>八		⋮	二	<—>	十六	
000	0	⋮	0000	0	1000	8
001	1	⋮	0001	1	1001	9
010	2	⋮	0010	2	1010	A
011	3	⋮	0011	3	1011	B
100	4	⋮	0100	4	1100	C
101	5	⋮	0101	5	1101	D
110	6	⋮	0110	6	1110	E
111	7	⋮	0111	7	1111	F

要求：迅速、准确，烂熟于心。

海纳百川 自强不息 厚德笃学 知行合一



## 2.1.2 数据的表示



### 1. 真值和机器数

➤ **真值**：数据的数值通常以正(+)、负(-)号后跟绝对值来表示，称之为“真值”。

➤ **机器数**：在计算机中正负号也需要数字化，一般用0表示正号，1表示负号。把符号数字化的数称为机器数。

- 逻辑型数据
- 字符型数据

- ✓ ASCII 码                      EBCDIC 码
- ✓ 字符串
- ✓ 汉字

- 数值型数据

- ✓ 定点小数              整数              浮点数
- ✓ 二—十进制数（BCD码）

海纳百川 自强不息 厚德笃学 知行合一



## 2. 逻辑型数据

➤ 逻辑型数据只有两个值：真 和 假，正好可以用二进制码的两个符号分别表示，

例如 1 表示 真

则 0 表示 假

- 不必使用另外的编码规则。
- 对逻辑型数据可以执行逻辑的 与 或 非等基本逻辑运算。其规则如下：

海纳百川 自强不息 厚德笃学 知行合一



# 逻辑型数据基本运算规则

中华人民共和国成立70周年

1949—2019  
大连理工大学 建校70周年

一个二进制数位可以用来表示一个二值逻辑型的变量，更准确的说法应该是一个基2码位而不是二进制的数位，因为逻辑型数据不存在进位关系（位间独立）。这里的与、或、非逻辑可以用与门、或门、非门电路实现。

X	Y	X与Y	X或Y	X的非
0	0	0	0	1
0	1	0	1	1
1	0	0	1	0
1	1	1	1	0

海纳百川 自强不息 厚德笃学 知行合一



### 3. 字符型数据的表示

中华人民共和国成立70周年 大连理工大学 建校70周年

➤ 字符作为人一机联系的媒介，是最重要的数据类型之一，当前的西文字符集由 **128** 个符号组成，通常用 **8** 位二进制编码，即用一个字节来表示一个字符。常用两个标准字符集：

**EBCDIC码: Extended Binary Coded  
Decimal Interchange Code**

**ASCII 码: American Standard Code for  
Information Interchange**

- **ASCII**码用**7**位二进制表示一个字符，总共**128**个字符元素，包括**10**个十进制数字（**0-9**）、**52**个英文字母（**A-Z**和**a-z**）、**34**个专用符号和**32**个控制符号。**ASCII**码字符集具体编码如下表所示：

海纳百川 自强不息 厚德笃学 知行合一



# ASCII 码字符集



L \ H	000	001	010	011	100	101	110	111
0000	NUL	DLE	SP	0	@	P	`	p
0001	SOH	DC1	!	1	A	Q	a	q
0010	STX	DC2	"	2	B	R	b	r
0011	ETX	DC3	#	3	C	S	c	s
0100	EOT	DC4	\$	4	D	T	d	t
0101	ENG	NAK	%	5	E	U	e	u
0110	ACK	SYN	&	6	F	V	f	v
0111	BEL	ETB	'	7	G	W	g	w
1000	BS	CAN	(	8	H	X	h	x
1001	HT	EM	)	9	I	Y	i	y
1010	LF	SUB	*	:	J	Z	j	z
1011	VT	ESC	+	;	K	[	k	{
1100	FF	FS	,	<	L	\	l	
1101	CR	GS	-	=	M	]	m	}
1110	SO	RS	.	>	N	↑	n	~
1111	SI	US	/	?	O	←	o	DEL

注：H 表示高 3 位，L 表示低 4 位。

海纳百川 自强不息 厚德笃学 知行合一





## 5. 汉字的表示与编码

中华人民共和国成立70周年 大连理工大学 建校70周年

➤ 通常用两个字节表示一个汉字。

➤ 为了与西文字符编码相区别（西文的ASCII码的最高一位编码值为 0），表示一个汉字时，把两个字节的最高一位的编码值设定为 1，则该编码集的最多编码数量为  $128 \times 128$ 。

➤ 这种编码方案与西文传送中的把 ASCII 码的最高一位用作奇偶校验位有矛盾。

海纳百川 自强不息 厚德笃学 知行合一



# 汉字的表示方法

中华人民共和国成立70周年 大连理工大学 建校70周年

## ➤ 汉字内码

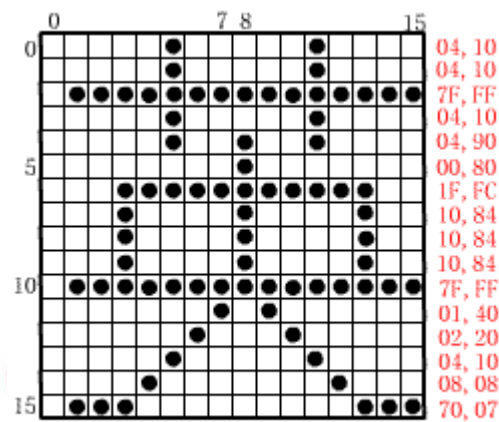
内码是用于汉字信息的存储、交换、检索等操作的机内代码，它一般采用两个字节的二进制形式表示一个汉字。《通讯用汉字字符集（基本集）及其交换码标准》，81年，6763个汉字，682个非汉字字符。

## ➤ 汉字输入编码

输入编码是为了使用西文标准键盘把汉字输入到计算机中，其编码方法主要有数字编码、拼音码和字形编码等。

## ➤ 汉字字模输出编码

字模编码是以点阵方式用来描述汉字字形的代码，它是汉字的输出形式。



# 汉字的表示

中华人民共和国成立70周年 大连理工大学 建校70周年

## (1) GB2312-80国标码

➤ 国标码：我国在1981年颁布了《通讯用汉字字符集（基本集）及其交换码标准》GB2312-80方案，简称国标码，共7445字，**各用两字节表示**。

- 各种字母、数字、符号682个
- 一级 3755个汉字，按拼音排序
- 二级 3008个汉字，按部首排列

海纳百川 自强不息 厚德笃学 知行合一



## (2) 区位码

- 将国标码中的字符按其位置划分成94个区(行)，每个区中94个字符(列)

中华人民共和国成立70周年

1949—2019  
大连理工大学 建校70周年

字符的区位码为其区号和位号

位号

区号

1 ..... 94

1  
~  
9区

字母、数字、各种符号等 682个

10  
~  
15区

空

16  
~  
55区

一级汉字 3755个

56  
~  
87区

二级汉字 3008个

88  
~  
94区

空

区位码是国标码的变形： 国标码=区位码+2020H



# 汉字在计算机内的表示

中华人民共和国成立70周年 大连理工大学 建校70周年

汉字内码：机内码，是指机器内部处理和存储汉字的一种代码。

常用的机内码在国标码基础上每个字节最高位置1。  
机内码=国标码+8080H=区位码+A0A0H

例：“京”字国标码为3E29H，其机内码为BEA9H，  
其区位码为1E09H。

海纳百川 自强不息 厚德笃学 知行合一



## (4) 汉字的存储

汉字的存储的含义：

汉字内码的存储

字形码的存储

中华人民共和国成立70周年

1949—2019  
大连理工大学 建校70周年

➤ **字形码**：也称字模码，目前计算机显示器和打印机都用点阵表示汉字字形代码，它是汉字的输出形式。

根据输出汉字的要求不同，点阵的多少也不同。简易型汉字为16X16点阵，提高型汉字为24X24点阵，甚至更高。

• **字模点阵**的信息量很大，需要占的存储空间大，点阵只能用来构成字库，而不能用于机内存储。字库中存储了每个汉字的点阵代码，当显示输出时才检索字库，输出字模点阵，得到字形

• **汉字内码**是用于汉字信息的存储、交换、检索等的机内代码，内码比字形点阵码占用空间少，一般用两个字节就可以表示一个汉字。

海纳百川 自强不息 厚德笃学 知行合一



## (5) 汉字的输出

➤ 汉字输出有打印输出和显示输出两种形式。汉字显示器多采用与图形显示兼容的光栅扫描显示器，采用点阵字型：16×16 32×32 48×48。

➤ 以这种方式输出的汉字是利用设备可以画点的图形方式实现的，因此，常称这种汉字为**图形汉字**。

➤ **汉字显示原理**：通过键盘输入的汉字编码，首先要经代码转换程序转换成汉字机内代码，转换时要用输入码到码表中检索机内码，得到两个字节的机内码，字形检索程序由机内码检索字模库，查出表示一个字形的点阵送显示缓存的指定位置，显示输出。

汉字显示输出过程：输入码 → 机内码 → 字形码 → 显示器

• 汉字的点阵码存放在硬件ROM中或软件字库中。

海纳百川 自强不息 厚德笃学 知行合一





王选

北京大学教授，中国科学院院士，中国工程院院士



本页是计算机激光汉字编辑排版系统的试排样张。

主持研制成功**汉字激光照排系统**（1979）并主持研制的**中国第一台计算机激光汉字照排系统原理性样机**（1981）：把一张报纸排版所需的时间从以前的4个小时降到**20分钟**。

引领了汉字印刷术的二次革命。

第一次：中国古代四大发明之一的活字印刷术。

告别了火与铅的时代，迎来了计算机与激光的时代

## 汉字信息处理

段版样张

1979年7月1日

由计算机总局主持，北京大学、新华社、山东省电子局、潍坊电子局、潍坊电子厂、杭州五〇二厂、天津红城厂等单位协作完成

计算机激光汉字编辑排版系统主体工程研制成功

### 汉字编辑排版系统的工作流程和软件

一、汉字编辑排版系统的工作流程  
汉字编辑排版系统的工作流程，是指从输入、编辑、排版到输出的全过程。其基本流程如下：  
1. 输入：将待排的文字输入计算机。  
2. 编辑：对输入的文字进行编辑、修改、删除、插入等操作。  
3. 排版：根据编辑好的文字，按照一定的格式进行排版。  
4. 输出：将排版好的文字输出到打印机或显示器上。

汉字编辑排版系统的工作流程，是指从输入、编辑、排版到输出的全过程。其基本流程如下：  
1. 输入：将待排的文字输入计算机。  
2. 编辑：对输入的文字进行编辑、修改、删除、插入等操作。  
3. 排版：根据编辑好的文字，按照一定的格式进行排版。  
4. 输出：将排版好的文字输出到打印机或显示器上。

### 滚筒式激光照排机的工作原理

滚筒式激光照排机的工作原理，是指通过激光束在滚筒上扫描，将文字信息记录在感光材料上的过程。其基本工作原理如下：  
1. 激光束的发射：由激光器发出激光束。  
2. 激光束的扫描：通过反射镜和透镜，使激光束在滚筒上进行扫描。  
3. 感光材料的曝光：激光束照射到感光材料上，使其发生光化学反应。  
4. 显影和定影：将曝光后的感光材料进行显影和定影处理，得到最终的排版结果。

滚筒式激光照排机的工作原理，是指通过激光束在滚筒上扫描，将文字信息记录在感光材料上的过程。其基本工作原理如下：  
1. 激光束的发射：由激光器发出激光束。  
2. 激光束的扫描：通过反射镜和透镜，使激光束在滚筒上进行扫描。  
3. 感光材料的曝光：激光束照射到感光材料上，使其发生光化学反应。  
4. 显影和定影：将曝光后的感光材料进行显影和定影处理，得到最终的排版结果。

### 第四代排字机

激光和计算机的结合排版系统和信息处理系统的突破

第四代排字机，是指将激光技术和计算机技术相结合的排版系统。其基本特点如下：  
1. 自动化程度高：通过计算机控制，实现排版过程的自动化。  
2. 排版效率高：相比传统排版方式，第四代排字机的排版效率大大提高。  
3. 排版质量好：通过激光技术，实现高质量的排版效果。  
4. 排版灵活性强：可以根据不同的需求，灵活调整排版格式。

第四代排字机，是指将激光技术和计算机技术相结合的排版系统。其基本特点如下：  
1. 自动化程度高：通过计算机控制，实现排版过程的自动化。  
2. 排版效率高：相比传统排版方式，第四代排字机的排版效率大大提高。  
3. 排版质量好：通过激光技术，实现高质量的排版效果。  
4. 排版灵活性强：可以根据不同的需求，灵活调整排版格式。

名称	研制成功的年代和国家	特点
第一代	手动排字机	西文：一九四八年前 汉字：传统，效率低
第二代	无铅式	西文：一九五〇年代 汉字：无铅，效率提高
第三代	照排机	西文：一九六〇年代 汉字：照排，效率提高
第四代	激光照排机	西文：一九七〇年代 汉字：激光，效率最高

## 6. 数值数据在计算机内的格式

中华人民共和国成立70周年

1949—2019  
大连理工大学 建校70周年

定点小数:  $N = N_s \mid N_{-1} N_{-2} \cdots N_{-n}$

定点整数:  $N = N_s N_{n-1} \cdots N_1 N_0$

浮点数:  $N = M_s E_s E_{m-1} \cdots E_1 E_0 M_{-1} M_{-2} \cdots M_{-n}$

基为 2

IEEE 标准: 阶码用移码 尾数用原码

	符号位	阶码位	尾数数码位	总位数
短浮点数:	1	8	23	32
长浮点数:	1	11	52	64
临时浮点数:	1	15	64	80

海纳百川 自强不息 厚德笃学 知行合一

# 数值范围和数据精度

## ➤ 数值范围

数值范围是指一种类型的数据所能表示的最大值和最小值。

## ➤ 数据精度

通常指实数所能给出的有效数字位数；对浮点数来说，精度不够会造成误差，误差大量积累会出问题。

## ➤ 机内处理

数值范围与数据精度概念不同。在计算机中，它们的值与用多少个二进制位表示某种类型的数据，以及怎么对这些位进行编码有关。



## 二 — 十进制码 (BCD码)

➤ **BCD码**：在计算机中采用**4位二进制码**对每个十进制数位进行编码，一个多位的十进制数被表示为这种编码的数串。**4位二进制码**有**16种**不同的组合，从中选出**10种**来表示十进制数位的**0~9**，用**0000**，**0001**， $\dots$ ，**1001**分别表示**0**，**1**， $\dots$ ，**9**，每个数位内部满足二进制规则，而数位之间满足十进制规则，故称这种编码为“以二进制编码的十进制 (**Binary Coded Decimal**，简称**BCD**) 码”。

➤ **BCD码加法运算的修正规则**：

- ✓ 如果两个一位**BCD**码相加之和小于或等于**9**，不需要修正；
- ✓ 如相加之和大于**9**或产生进位，要进行加**6**修正，如果有进位，要向高位进位。



# BCD码



- BCD码有多种不同的编码方案： 例如：**8421码**，**84-2-1码**，余**3码**，格雷码 .....
- 可以进一步分为有权码和无权码：
  - ✓ 有权码：每位上的**1**代表确定的值
  - ✓ 无权码：无法确定每位上的**1**代表的值

海纳百川 自强不息 厚德笃学 知行合一



# 有权码      无权码



中华人民共和国成立70周年



1949 — 2019  
大连理工大学 建校70周年

	8421	84-2-1	余3码	格雷码
0	0000	0000	0011	0000
1	0001	0111	0100	0001
2	0010	0110	0101	0011
3	0011	0101	0110	0010
4	0100	0100	0111	0110
5	0101	1011	1000	1110
6	0110	1010	1001	1010
7	0111	1001	1010	1000
8	1000	1000	1011	1100
9	1001	1111	1100	0100

海纳百川 自强不息 厚德笃学 知行合一





# 如何判定码权

0 0000 ✓  
1 0111 ✓  $4 + (-2) + (-1)$

2 0110 ✓  $4 + (-2)$

3 0101 ✓  $4 + (-1)$

4 0①00 → ④

5 1011 ✓  $8 + (-2) + (-1)$

6 10①0 → ②

7 100① → ①

8 ①000 → ⑧

9 1111 ✓  $8 + 4 + (-2) + (-1)$

要验证每个码的值

从一编码求码权

结论

证明此编码系统为有权码

海纳百川 自强不息 厚德笃学 知行合一



# 如何判定码权

0 0011 ~~2+1 ≠ 0~~

1 0**1**00 → ①

2 010**1** → ①

3 01**1**0 → ②

4 0111

5 1000

6 1001

7 1010

8 1011

9 1100

验证各码的值  
从一编码求码权

结论

证明此编码系统为无权码

海纳百川 自强不息 厚德笃学 知行合一

## 2.1.3 校验码



要提高计算机的可靠性，除了采取选用更高可靠性的器件，更好的生产工艺等措施之外，还可以针对薄弱环节，从数据编码上想一些办法

即采用少量冗余的线路，在原有数据位之外再增加一到几个校验位，使新得到的由数据位和校验位构成的码字带上某种特性

在经过薄弱环节之后，则通过检查该码字是否仍保持有这一特性，来判断码字中的某一、二位的值是否发生了变化，即是否出现了错误，甚至于定位错误后，自动改正这一错误，这就是**检错纠错编码技术或校验码方法**

海纳百川 自强不息 厚德载物 学无止境



# 校验码

中华人民共和国成立70周年 1949—2019 大连理工大学 建校70周年

➤ **数据校验码**是一种常用的带有发现某些错误或自动改错能力的**数据编码方法**。其**实现原理**，是加进一些冗余码，使合法数据编码出现某些错误时，就成为非法编码。

- 通过检测编码的合法性来达到发现错误的目的。
- 合理地安排非法编码数量和编码规则：可以提高发现错误的能力，或达到自动改正错误的目的。

➤ **码距**：指任意两个合法码之间至少有几个二进制位不相同。仅有一位不同，称其码距为1。奇偶校验码为2。

- 合理增大码距，能提高发现错误的能力
- 但表示一定数量的合法码所使用的二进制位数要变多，增加了电子线路的复杂性和数据存储、数据传送的数量。



# 几种常用的校验码

中华人民共和国成立70周年 大连理工大学 建校70周年

介绍三种常用的校验码：

**奇偶校验码：**用于串、并行数据**传送中**

**循环冗余校验码：**用于串行数据**传送中**

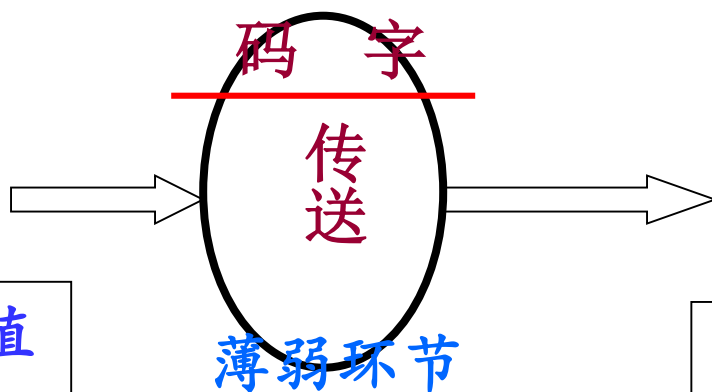
**海明校验码：**用于**并行**数据**传送中**（不在考试范围）

**运行过程的3步曲：**

原始数据

编码过程

形成校验位的值  
加进特征



结果数据

译码过程

检查收到的码字

发现 / 改正错误

# 1. 奇偶校验码

中华人民共和国成立70周年 大连理工大学 建校70周年

➤ **实现原理**：是使码距由1增加到2。若编码中有1位二进制数出错了，出错的编码就成为非法编码，就可以知道出现了错误。在原有的编码之上再增加一位校验位，原编码 $k$ 位，形成新的编码为 $k+1$ 位。增加的方法有2种：

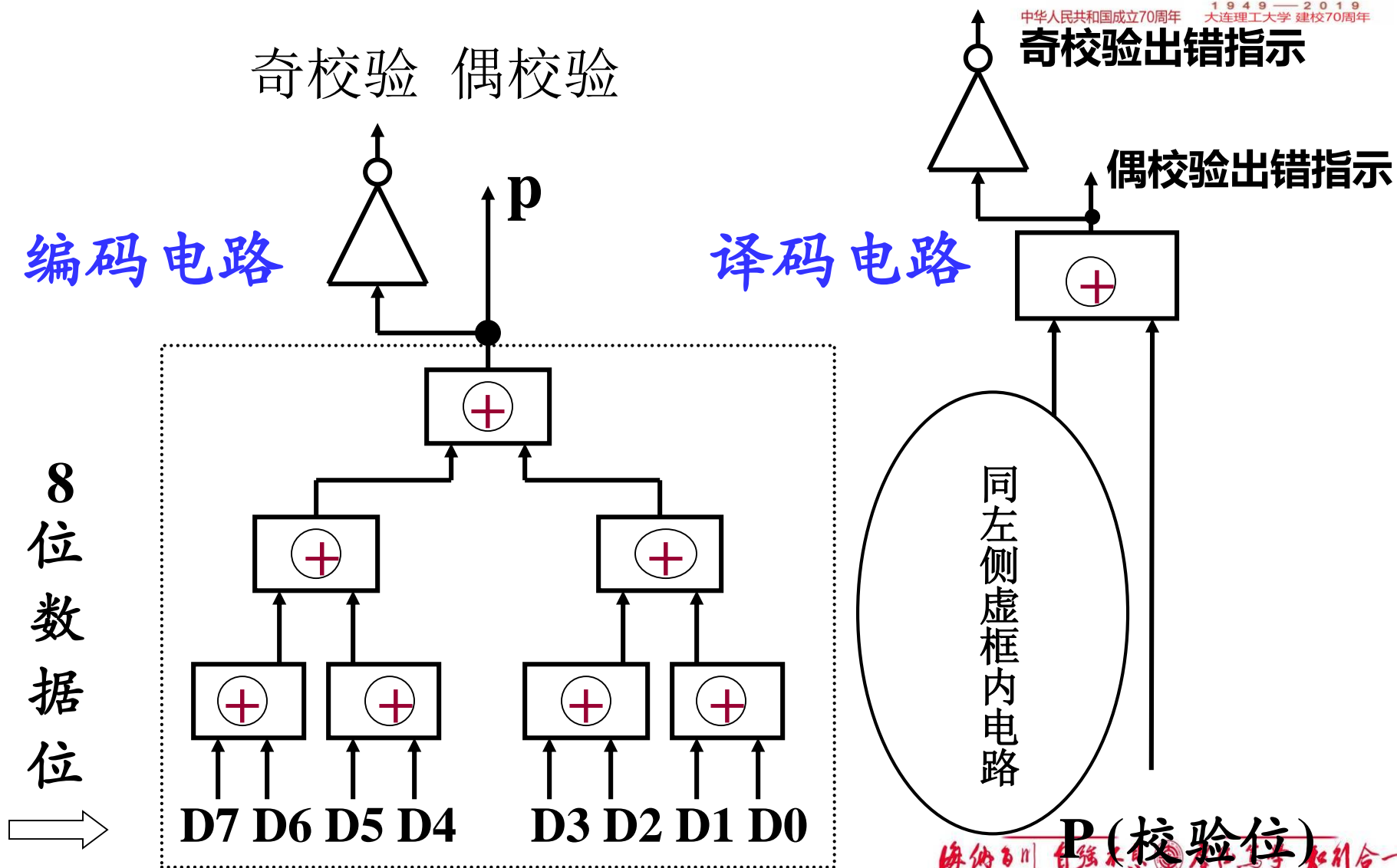
- **奇校验**：增加位的0或1要保证整个编码中1的个数为奇数个
- **偶校验**：增加位的0或1要保证整个编码中1的个数为偶数个

海纳百川 自强不息 厚德笃学 知行合一

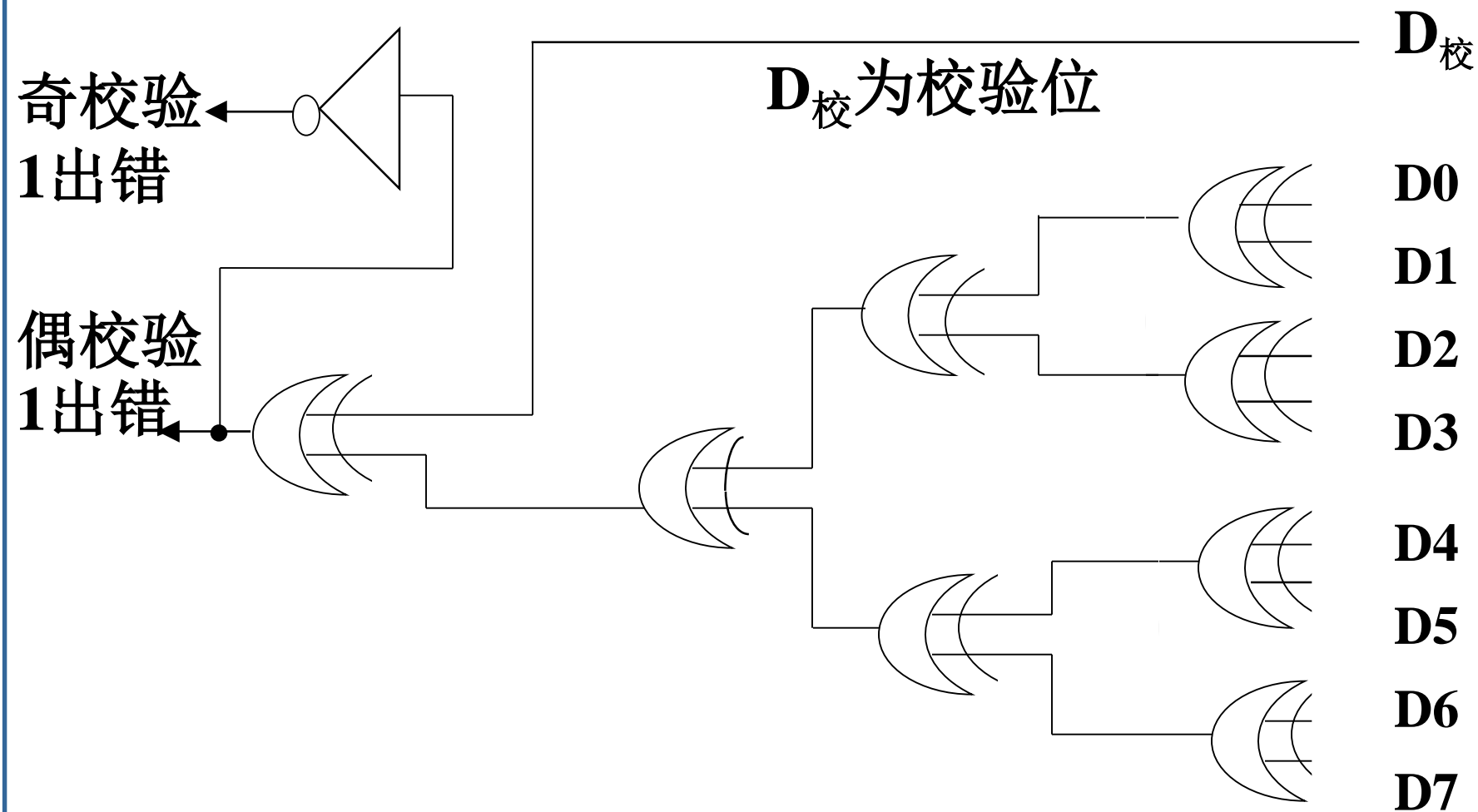


# 奇偶校验码的实现电路

中华人民共和国成立70周年 1949—2019  
大连理工大学 建校70周年



# 8位数据的奇偶校验码形成电路及检码电路



海纳百川 自强不息



【例】选择题：如果需要对3个字节的数据进行纵向奇偶校验，可再增加1个字节作为纵向校验字节，纵向校验的奇偶性与横向校验的奇偶性相同。若被校验的3个字节数据分别是A9H、B8H和CFH，则新增加的校验字节数据是（ ）。

A. AEH      B. DEH      C. D6H      D. 5EH

【分析】3个横向校验码A9H、B8H和CFH中1的个数均为偶数个，说明系统采用的是偶校验，4个字节（3个数据字节加上纵向校验字节）对应位中1的个数也应当为偶数个。在4个选项中只有B选项符合条件。

【归纳总结】计算机在进行大量字节（数据块）传送时，不仅每一个字节有一个奇偶校验位做横向校验，而且全部字节的同一位也设置一个奇偶校验位做纵向校验，这种横向、纵向同时校验的方法称为交叉校验。

【解题技巧】首先根据3个字节数据确定校验规则的奇偶性，增加的纵向校验字节要保证纵向每一位也符合同样的奇偶性。

答案：B

海纳百川 自强不息 厚德笃学 知行合一



## 2. 循环冗余校验码 (CRC)



- **CRC (Cyclical Redundancy Check)** 校验码一般是指 $k$ 位信息之后拼接 $r$ 位校验码。用于多位串行数据传送中的检错、纠错处理。
- **关键问题:**
  - ✓ 如何从 $k$ 位信息得到 $r$ 位校验码。
  - ✓ 如何从 $k+r$ 位码字判断是否出错，哪位出错。
- **实现过程:**
  - 在  $k$  位数据位串行移位输出的过程中，用带有异或门控制的移位寄存器形成  $r$  个校验位的值，跟随在数据位之后传送走。
  - 在接收端再对  $k+r$  位的码字进行合法与出错检查，若可能则自动改错。



# CRC码的编码方法

中华人民共和国成立70周年 大连理工大学 建校70周年

➤CRC整个编码长度为  $n=k+r$  位，故CRC码又叫  $(n, k)$  码。其编码方法如下：

- 假设被传送的  $k$  位二进制信息位用  $C(x)$  表示，系统选定的生成多项式用  $G(x)$  表示，将  $C(x)$  左移  $G(x)$  的最高次幂(即等于需要添加的校验位的位数  $r$ )，写作  $C(x) \cdot 2^r$

- 然后将  $C(x) \cdot 2^r$  除以生成多项式  $G(x)$ ，所得商用  $Q(x)$  表示，余数用  $R(x)$  表示。则有：

$$C(x) \cdot 2^r / G(x) = Q(x) + R(x) / G(x)$$

- 两边同时乘以  $G(x)$  并左移  $R(x)$  得到：

$$C(x) \cdot 2^r - R(x) = Q(x) \cdot G(x)$$

感谢 | 自强不息 厚德与学 知行合一



➤**CRC**编码采用的是按位加、减法，即不考虑进位与借位，运算规则为： $0 \pm 0=0$ ， $0 \pm 1=1$ ， $1 \pm 0=1$ ， $1 \pm 1=0$

故有：

$$C(x) \cdot 2^r + R(x) = Q(x) \cdot G(x)$$

上式中，等式左边即为所求的n位CRC码，其中余数表达式R(x)就是校验位(r位)。且等式两边都是G(x)的倍数。

- 发送信息时将等式左边生成的n位**CRC**码送给对方。当接收方接到n位编码后，同样除以**G(x)**，如果传输正确则余数为**0**，否则，可以根据余数的数值确定是哪位数据出错。



【例】有一个 (7, 4) 码 (即CRC码为7位, 信息码为4位)

已确定生成多项式为:  $G(X) = X^3 + X + 1 = 1011$

被传输的信息  $C(x) = 1001$ , 求  $C(x)$  的CRC码。

解:  $C(x)$  左移  $r = n - k = 3$  位

$$\text{即 } C(x) \cdot 2^r = 1001 \times 2^3 = 1001000$$

将上式模2采用除法 除以给定的  $G(x) = 1011$ :

$$1001000 / 1011 = 1010 + 110 / 1011$$

得到余数表达式:  $R(x) = 110$

所求CRC码:

$$C(x) \cdot 2^3 + R(x) = 1001000 + 110 = 1001110$$





中华人民共和国成立70周年



1949 — 2019  
大连理工大学 建校70周年

$$\begin{array}{r} 1010 \\ 1011 \overline{) 1001000} \\ \underline{1011} \phantom{000} \\ 1000 \\ \phantom{10} \underline{1011} \\ 110 \end{array}$$

(A)

$$\begin{array}{r} 1010 \\ 1011 \overline{) 1001111} \\ \underline{1011} \phantom{000} \\ 1011 \\ \phantom{10} \underline{1011} \\ 1 \end{array}$$

海纳百川 自强不息 厚德笃学 知行合一 (B)



# CRC码的查错表



CRC	A7	A6	A5	A4	A3	A2	A1	余数	出错
正确	1	0	0	1	1	1	0	000	
某一位出错	1	0	0	1	1	1	1	001	A1
	1	0	0	1	1	0	0	010	A2
	1	0	0	1	0	1	0	100	A3
	1	0	0	0	1	1	0	011	A4
	1	0	1	1	1	1	0	110	A5
	1	1	0	1	1	1	0	111	A6
	0	0	0	1	1	1	0	101	A7

- 收到的CRC码除以约定的生成多项式 $G(x)$ ，如果余数为0则传输无误，否则传输错误，根据所得余数值就可找出错误并取反纠正。
- 上表详细说明了CRC码1001110在传送时某一位出错后的判断与纠正方法 [ $C(X) = 1001$ 、 $G(x) = 1011$ ]。

海纳百川 自强不息 厚德笃学 知行合一





# 生成多项式 $G(x)$ 的确定



$G(x)$ 是一个约定的除数，用来产生校验码。从检错和纠错的要求出发，它并不是随意选择的，为了得到 $r$ 位余数， $G(x)$ 必须是 $r+1$ 位。

应满足下列要求：

- ✓任何一位发生错误都应使余数不为0；
- ✓不同位发生错误应使余数不同；
- ✓余数继续模2除，应使余数循环。

海纳百川 自强不息 厚德笃学 知行合一



# CRC的译码与纠错



中华人民共和国成立70周年 大连理工大学 建校70周年

- 将收到的循环校验码用约定的生成多项式 $G(x)$ 去除，如果码字无误则余数应为0, 如有某一位出错，则余数不为0, 不同位数出错余数不同。

CRC	A7	A6	A5	A4	A3	A2	A1	余数	出错
正确	1	0	0	1	1	1	0	000	
某一位出错	1	0	0	1	1	1	1	001	A1
	1	0	0	1	1	0	0	010	A2
	1	0	0	1	0	1	0	100	A3
	1	0	0	0	1	1	0	011	A4
	1	0	1	1	1	1	0	110	A5
	1	1	0	1	1	1	0	111	A6
	0	0	0	1	1	1	0	101	A7

海纳百川 自强不息 厚德笃学 知行合一



# 校验码小结



- (1) **k 位码**有  $2^k$  个编码状态, 全用于表示合法码, 则任何一位出错, 均会变成另一个合法码, 不具有检错能力。
- (2) 从一个合法码变成另一个合法码, 至少要改变几位码的值, 称为最小码距 (码距), 码距和编码方案将决定其检错纠错能力。

奇偶校验码的码距为 2,

常用循环冗余码的码距 $r+1$ , 例如 $r=3$ , 其码距为4。

海纳百川 自强不息 厚德笃学 知行合一



# 校验码小结



中华人民共和国成立70周年



1949 — 2019  
大连理工大学 建校70周年

(3)  **$k+1$  位码**，只用其  $2^k$  个状态，可以使码距为 2，  
如果一个合法码中的一位错了，就成为非法码，  
通过检查码字的合法性，就**得到检错能力**，这就是奇偶校验码，只能发现1位错，不具备纠错能力。

**$k+1$  位码时，有  $2^k$  个奇数（ $2^k$  个奇状态）；  
有  $2^k$  个偶数（ $2^k$  个偶状态）。**

海纳百川 自强不息 厚德笃学 知行合一

