



## 《智能思维与计算社会科学方法论》

第一单元：数据——计算的基础资源

主讲教师：郑大庆（博士、副教授）



## 智能思维与计算社会科学方法论：数据观

- 第一讲：数据存储形式
- 第二讲：数据组织和管理形式-1
- 第三讲：数据组织和管理形式-2
- 第四讲：Python数据访问实验
- 第五讲：数据库与数据可视化分析



## 《智能思维与计算社会科学方法论》

第一讲 数据存储与组织

主讲教师：郑大庆（博士、副教授）



## 本讲主要内容

- 理解0和1的思维
  - ✓ 语义符号化的典型案例
  - ✓ 思维方式与逻辑运算
  - ✓ 二进制与算术元素
  - ✓ 编码与符号运算

## 2024年开年王炸：AI辅助生成《西游记》



## 理解0和1的思维



### 语义符号化

✓语义 (Semantic)：语言所蕴含的意义。符号是语言的载体，符号本身没有任何意义，只有被赋予含义的符号才能够被使用，这时候语言就转化为信息，而语言的含义就是语义。

## 理解0和1的思维

### 语义符号化的典型案例

- ✓语义符号化：将现实世界的语义用符号表达，进而进行基于符号计算的一种思维。
  - 将语义表达为不同的符号，便可以采用不同的工具进行计算；
  - 将符号赋予不同语义，则能计算不同的现实世界问题；

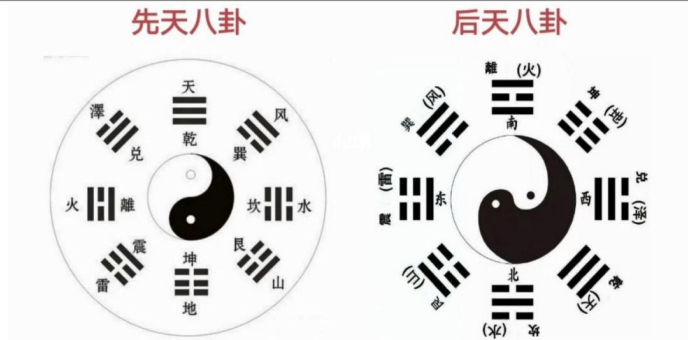


## 易经的语义符号化

Diagram illustrating the Bagua (Eight Trigrams) and their semantic applications. The central Taijitu is surrounded by the eight trigrams: 乾 (Qian), 坤 (Kun), 坎 (Kan), 离 (Li), 震 (Zhen), 巽 (Xun), 艮 (Gen), and 兑 (Dui). The text explains the semantic applications of these trigrams in three contexts: natural phenomena, family space, and animal space.

- 本义 (自然现象)：乾为天，坤为地，坎为月为水，离为日为火，震为雷为动，巽(xùn)为风为入，艮(gěn)为山为止，兑为泽为悦；
- 用1 (家庭空间)：乾为父，坤为弱，坎为中男，离为中女，震为长男，巽为长女，艮为少男，兑为少女；
- 用2 (身体空间)：乾为首，坤为腹，坎为耳，离为目，震为足，巽为股，艮为手，兑为口；
- 用3 (动物空间)：乾为马，坤为牛，坎为豕，离为雉，震为龙，巽为鸡，艮为狗，兑为羊；

## 先天八卦与后天八卦



## 语义符号化

### 一个理解与抽象的过程

- ✓对现实世界现象的深入理解，抽象出普适的概念（或称为本体），进而可将概念符号化，进行各种计算；
- ✓再将符号赋予不同的语义，可以处理不同语境的问题。

## 思维方式与逻辑思维

➤逻辑：事物因果之间所遵循的规律，是现实中普适的思维方式；

➤逻辑的具体表现形式是命题和推理

✓命题：语句的涵义，由一语句表达的内容为“真”或为“假”的一个判断；

- 命题1：“罗素不是一位小说家”
- 命题2：“罗素是哲学家”

✓推理：依据由简单命题的判断推导出复杂命题的判断结论的过程。

- 复杂命题的推理可以被认为是关于命题的一组逻辑运算的过程。

## 思维方式与逻辑思维

➤复杂命题的推理可以被认为是关于命题的一组逻辑运算的过程。

✓与（AND）、或（OR）、非（NOT）、异或（XOR）

$\begin{array}{r} 0 \\ \text{AND} \\ 0 \\ \hline 0 \end{array}$	$\begin{array}{r} 0 \\ \text{AND} \\ 1 \\ \hline 0 \end{array}$	$\begin{array}{r} 1 \\ \text{AND} \\ 0 \\ \hline 0 \end{array}$	$\begin{array}{r} 1 \\ \text{AND} \\ 1 \\ \hline 1 \end{array}$	注： 1 表 真 0 表 假
$\begin{array}{r} 0 \\ \text{OR} \\ 0 \\ \hline 0 \end{array}$	$\begin{array}{r} 0 \\ \text{OR} \\ 1 \\ \hline 1 \end{array}$	$\begin{array}{r} 1 \\ \text{OR} \\ 0 \\ \hline 1 \end{array}$	$\begin{array}{r} 1 \\ \text{OR} \\ 1 \\ \hline 1 \end{array}$	
$\begin{array}{r} 0 \\ \text{NOT} \\ 1 \\ \hline 1 \end{array}$	$\begin{array}{r} 1 \\ \text{NOT} \\ 0 \\ \hline 1 \end{array}$			
$\begin{array}{r} 0 \\ \text{XOR} \\ 0 \\ \hline 0 \end{array}$	$\begin{array}{r} 0 \\ \text{XOR} \\ 1 \\ \hline 1 \end{array}$	$\begin{array}{r} 1 \\ \text{XOR} \\ 1 \\ \hline 0 \end{array}$	$\begin{array}{r} 1 \\ \text{XOR} \\ 0 \\ \hline 1 \end{array}$	

## 形式逻辑的关键进展

- 亚里士多德（公元前384-322）：提出关于逻辑的一些基本规律；
- 莱布尼兹（1646-1716）：把形式逻辑符号化，对人的思维进行运算和推理；
- 布尔（1815-1864）：提出了布尔代数，一种基于二进制逻辑的代数系统；



## 二进制与算术运算

### ➤ 数值表示

- ✓ 数码和带有权值的数位；
  - 数码：**r进制数码**
  - 数位： 在一个数值中的位置
    - r进制数位的权值为r的幂次方，表示逢r进1，借1当r。
  - 例题： $(1011110001.01011)_2$



## 二进制与算术运算

### ➤ 二进制的优势

- ✓ 二进制算术运算规则简单
- ✓ 二进制算术运算和逻辑运算可以统一
- ✓ 表示2种状态的电子元器件相对容易（开关、灯泡、二极管）



## 二进制与算术运算

### ➤ 符号表示

- ✓ 二进制数的符号可以用0和1表示；
- ✓ 机器数可用原码、反码和补码表示，不同的表示方法有不同的计算规则。

### ➤ 用补码把减法转化成加法

真实数值 (带符号的n位 二进制数)	十进制数	机器数 (n+1位二进制数, 其中第n+1位表符号, 0表示正号, 1表示符号)		
		原码	反码	补码
+11...11	+ (2 <sup>n</sup> -1)	0 11...11	0 11...11	0 11...11
+10...00	+2 <sup>n-1</sup>	0 10...00	0 10...00	0 10...00
+00...00	+0	0 00...00	0 00...00	0 00...00
-00...00	-0	1 00...00	1 11...11	0 00...00
-10...00	-2 <sup>n-1</sup>	1 10...00	1 01...11	1 10...00
-11...11	-(2 <sup>n</sup> -1)	1 11...11	1 00...00	1 00...01
-100...00	-2 <sup>n</sup>			1 00...00
正数的原码、反码同补码形式是一样的，最高位为0表示正数				
说明		负数的最高位为1，表示负数。其余同真实数值的二进制数	负数的最高位为1，表示负数。其余在反码基础上最低位加1后形成的。它的二进制数基础上逐位取反	负数的最高位为1，表示负数。其余在反码基础上最低位加1后形成的。它的负数不包括0，但包括-2 <sup>n</sup>
机器数由于受到表示数值的位数的限制，只能表示一定范围内的数，超出此范围则为“溢出”				

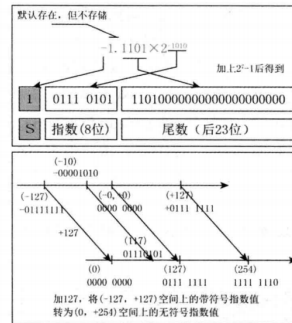




## 二进制与算术运算

### 小数点表示

- S** (全为小数)  
定点数，小数点位置固定（默认在符号位S的后面）
- S** (全为整数)  
定点数，小数点位置固定（默认在尾部）
- S** 指数(8位) 尾数(后23位)  
浮点数，32位表示单精度数（相当于科学计数法 $1.x \times 2^y$ ）  
S为符号位，x为23位尾数，y为8位指数
- S** 指数(11位) 尾数(后52位)  
浮点数，64位表示双精度数（相当于科学计数法 $1.x \times 2^y$ ）  
S为符号位，x为52位尾数，y为11位指数



## 编码与符号运算

### 编码的概念

✓以若干位数码或符号的不同组合来表示数值型信息的方法，是人为地将若干数码或符号的每种组合指定一种唯一的含义。

### 编码的主要特征

- ✓唯一性：确定、唯一的含义
- ✓公共性：相关者都认同、遵守和使用这种编码
- ✓规律性：有一定的规律和一定的编码规则

## 编码与符号运算

### 用0,1组合编码字母与符号：ASCII

✓ASCII (American Standard Code for Information Interchange) 用7位二进制数表示一个常用符号的一种编码

- “We are students” 对应的存储

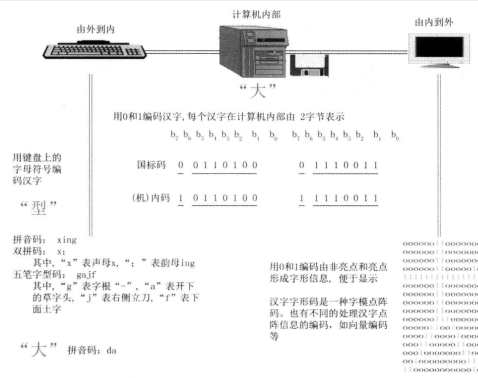
b <sub>7</sub> b <sub>6</sub> b <sub>5</sub> b <sub>4</sub> b <sub>3</sub> b <sub>2</sub> b <sub>1</sub> b <sub>0</sub>	000	001	010	011	100	101	110	111
0000	NUL	DLE	SP	0	@	P	`	P
0001	SOH	DC1	!	1	A	Q	a	Q
0010	STX	DC2	+	2	B	R	b	R
0011	ETX	DC3	#	3	C	S	c	S
0100	EOT	DC4	\$	4	D	T	d	T
0101	ENQ	NAK	%	5	E	U	e	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

## 编码与符号运算

### 用字母符号组合编码汉字

- ✓国标码：1981年，《通讯用汉字字符集（基本集）及其交换标准》（GB2312-1980方案）
  - 由2个字节表示一个汉字的编码，每个自己的最高位为0；
- ✓汉字机内码（内码）：对应国标码基础上将每字节的最高位设为1。汉字内码是用最高位均为1的2个字节表示一个汉字，是计算机内部处理、存储汉字信息所使用的统一编码。

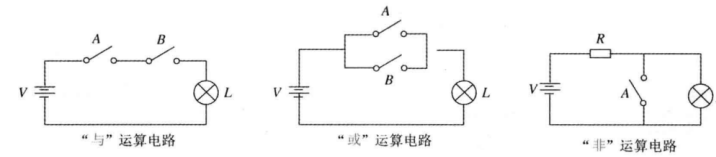
## 汉字编码



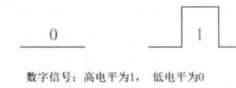
## 0和1与电子元器件

基本的逻辑运算可由开关及其电路连接来实现。

电路接通为1, 电路断开为0



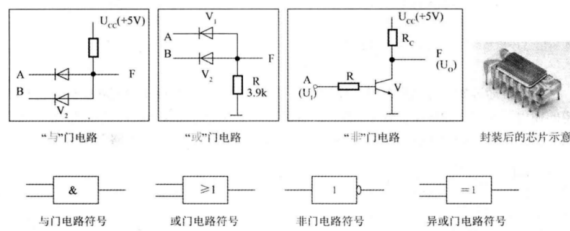
二极管高电平为1, 低电平为0



## 0和1与电子元器件

基本的逻辑运算可由开关及其电路连接来实现

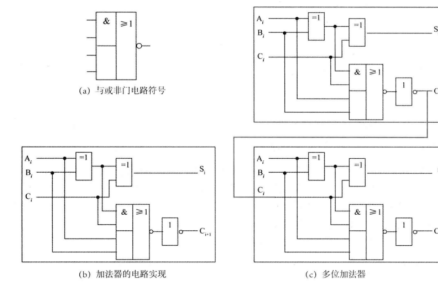
用二极管实现的基本集成电路示意图



## 0和1与电子元器件

基本的逻辑运算可由开关及其电路连接来实现

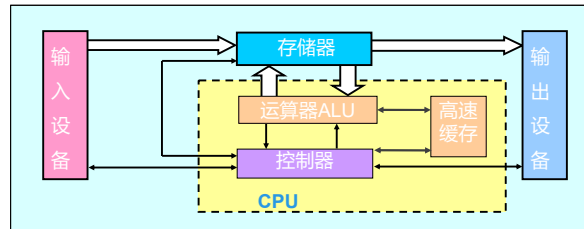
加法器的构建



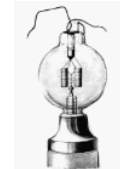
## 冯·诺依曼计算机

冯·诺依曼做出了现代意义上的计算机，对现代计算机的发展产生重要影响，被称为“计算机之父”，现在的计算机也被称为“冯·诺依曼机”

✓ 存储程序的思想

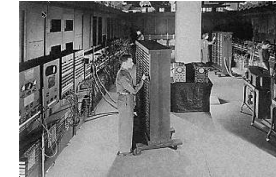


## 电子管时代的计算机



人类第一只电子管  
(真空二极管), 1895

存储0和1的元器件

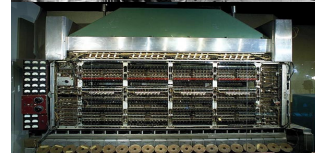
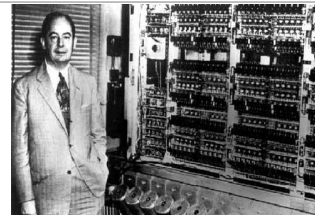


电子管计算机ENIAC, 1946年, 17468只电子管



## 电子管时代的计算机

- 冯·诺伊曼 (Von Neumann) 电子计算机EDVAC问世
- 将运算和存储分离，运算速度却比拥有18000个电子管的“ENIAC”提高了10倍
- 结构上的创新：“冯·诺伊曼计算机”。



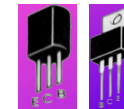
?

运算规则和数据是否可事先存储于存储器中，以便机器连续的执行呢？  
运算和存储怎样分离呢？

## 晶体管时代的计算机



人类第一只晶体管  
(真空二极管), 1947



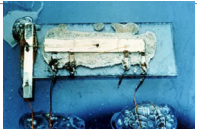
?

怎样使体积更小？可靠性更高？可控性更灵活？



第一台晶体管计算机TRADIC, 1953

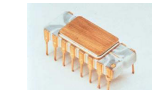
## 集成电路时代的计算机



集成电路的发明,1959



J. Kilby, 集成电路发明者



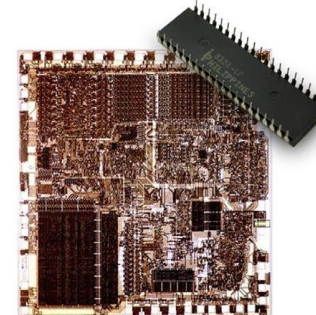
封装后的集成电路芯片



第三代计算机IBM360,1964

能否将复杂的电路封装后作为新电路设计的元件呢?  
复杂的电路 → 集成 → 封装 → 应用?

## 超大规模集成电路(VLSI)时代的计算机



VLSI芯片及其封装的内部电路

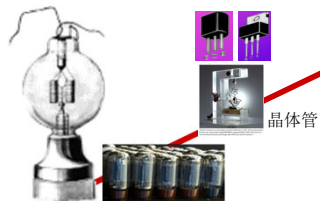
摩尔定律---每18个月芯片能力增长一倍



第四代计算机—个人计算机,1981

## 计算机元器件的发展

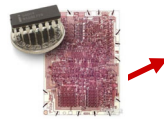
从表示-自动存储-自动执行的角度



电子管:可自动控制0和1变化的元件



晶体管



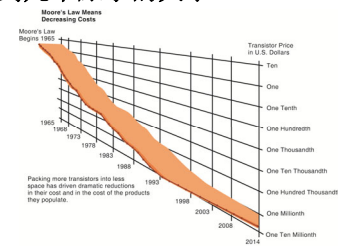
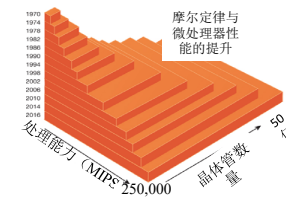
超大规模集成电路(VLSI)

集成电路:可自动实现一定变换的元件

- 体积越来越小;
- 可靠性越来越高;
- 电路规模越来越大;
- 速度越来越快;
- 功能越来越强大;

## 摩尔定律和微处理能力

- 计算力每18个月增加一倍;
- 纳米技术: 可以将晶体管缩小到几个原子的大小



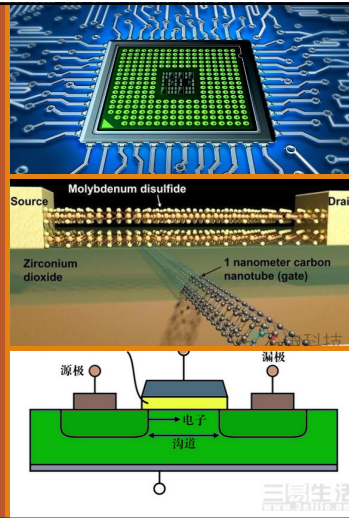
# 纳米 (nm) 到底在讲什么?

随着半导体制造技术的进步，芯片越做越小，带来的是芯片集成度与性能不断提升，功耗越来越低，但同时也意味着技术含量和成本越来越高。

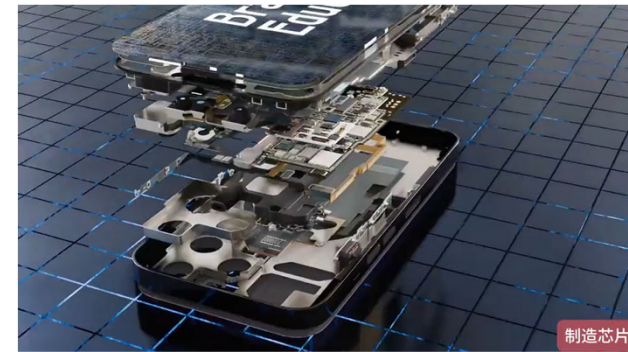
对于中国大陆芯片制造公司来说，尚无法造出代表目前芯片制造领域最高点的7nm、5nm芯片，但28nm以上成熟工艺领域的芯片制造，则是满满的机会。

虽然28nm目前是先进工艺和成熟工艺的分界点，但这个分界点并不固定。如果大批代工厂都突破了14nm，那14nm就成为了成熟工艺。

\* 制程：晶体管中，源极与漏极之间的距离；



# 芯片是如何研制的?



# 重磅：中国研究团队实现亚1纳米栅极长度晶体管

2022年3月12日，据清华大学官方宣布，清华大学集成电路学院任天令教授团队在小尺寸晶体管研究方面取得重大突破，首次实现了具有亚1纳米栅极长度的晶体管，并具有良好的电学性能。

nature

Explore content About the journal Publish with us Subscribe

Article 18 March 2022

## Vertical MoS<sub>2</sub> transistors with sub-1 nm gate lengths

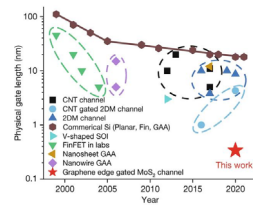
Fan Wu, He Tian, Yang Shen, Zhan Zhou, He Ben, Guangyuan Guo, Yubin Sun, Yi Xiang & Tian Ling Ren

Nature 603, 259–264 (2022) | Cite this article

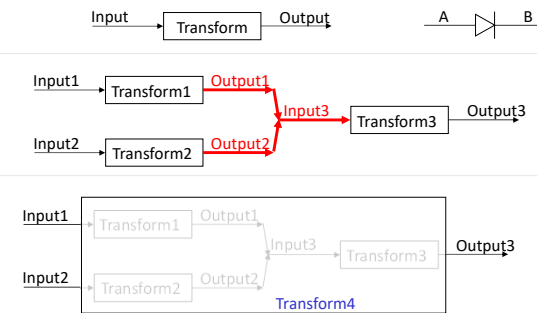
7376 Accession | 72 Abstracts | Metrics

### Abstract

Ultra-scaled transistors are of interest in the development of next-generation electronic devices<sup>1–3</sup>. Although atomically thin molybdenum disulfide (MoS<sub>2</sub>) transistors have been reported<sup>4</sup>, the fabrication of devices with gate lengths below 1 nm has been challenging<sup>5</sup>. Here we demonstrate side-wall MoS<sub>2</sub> transistors with an atomically thin channel and a physical gate length of sub-1 nm using the edge of a graphene layer as the gate electrode. The approach uses large-area graphene and MoS<sub>2</sub> films grown by chemical vapour deposition for

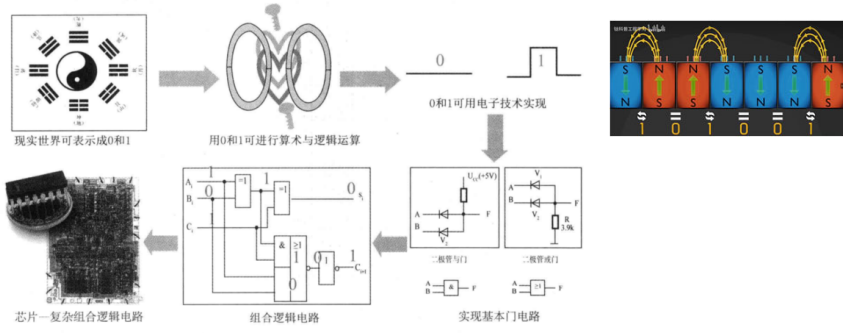


# 计算机的元器件发展





# 本章小结



谢谢!  
Questions & Comments