

第3章

存储子系统

SoC 的存储子系统由寄存器、缓存、内存（主存）和外存（辅存）四级结构构成，用于存储程序和各种数据，如图 3.1 所示。

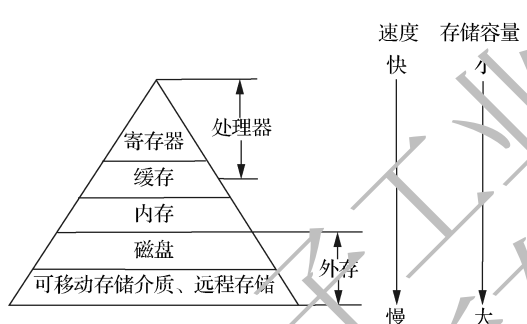


图 3.1 存储子系统的层级结构

图 3.2 所示为典型处理器的多级存储子系统。包括处理器内部的一级缓存和二级缓存，由 SRAM、DRAM 和 ROM 等构成的内存，以及由存储卡和磁盘等构成的外存。处理器操作的是寄存器中的数据，寄存器能够以接近处理器的时钟频率工作，当在寄存器中没有找到数据时，则在一级缓存中查找，并将查找到的数据加载至寄存器，如果一级缓存中也没有，则在二级缓存中查找，依次类推。

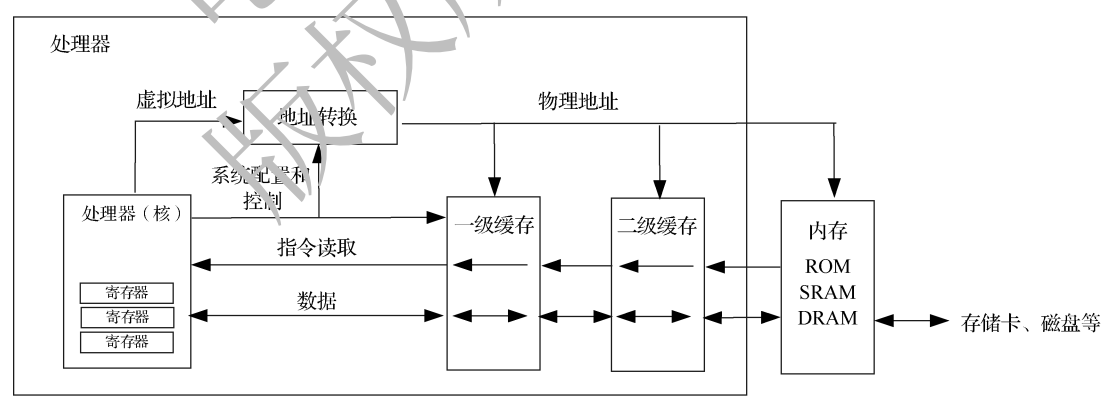


图 3.2 典型处理器的多级存储子系统

本章首先介绍存储器，然后讨论存储子系统的层次，最后介绍 DRAM 和闪存。

3.1 存储器

在数字系统中，存储器用于存储数据和程序，按存储介质不同，可分为磁性存储器、光存储器和半导体存储器；按存取方式不同，可分为 RAM、串行存取存储器、相联存储器等。

(1) RAM。

RAM 中的存储单元可随意存取，且存取时间与存储单元的物理位置无关。

(2) 串行存取存储器。

串行存取存储器中的存储单元只能按某种顺序存取，存取时间与存储单元的物理位置有关。移位寄存器和 FIFO 存储器都属于此类。

(3) 相联存储器。

相联存储器的存储单元不按地址而按内容来存取，可以实现快速查找快表，又称关联存储器或按内容寻址存储器。

3.1.1 半导体存储器

半导体存储器是一种以半导体电路作为存储媒介的存储器，包括 ROM 和 RAM，如图 3.3 所示。

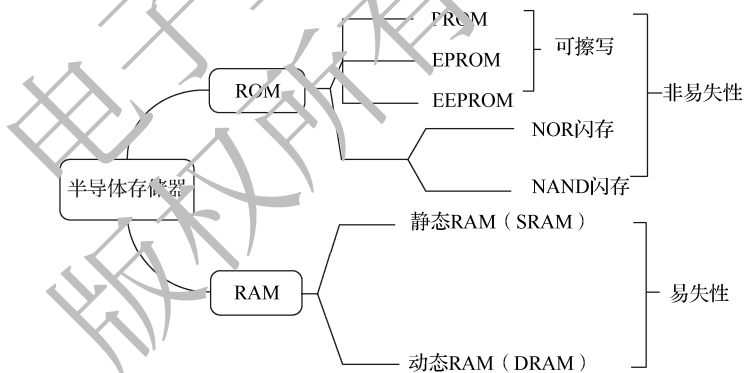


图 3.3 半导体存储器类型

(1) ROM。

ROM 是非易失性存储器，用于存储程序和静态数据，即使断电，数据也不会丢失。

ROM 意为只读存储器。所谓只读，是相对于处理器的读/写控制电平和控制逻辑而言的。有些 ROM 可读可写，只不过需要专门的写、擦除电平和时序，但是整体上都被称为 ROM。

根据重新编程（写）的次数和机制，ROM 可分为以下 4 种。



- ① PROM (Programmable ROM, 可编程 ROM): 只能被编程一次。
- ② EPROM (Erasable Programmable ROM, 可擦可编程 ROM): 可擦写达 1000 次。
- ③ EEPROM (Electrically-Erasable Programmable ROM, 电擦除可编程 ROM)。
- ④ 闪存 (Flash Memory): 不仅可读可写, 还具备非易失性。

(2) RAM。

RAM 是易失性存储器, 用于暂时存储程序和数据, 断电后数据会丢失。

RAM 在使用过程中, 既可利用程序随时写入信息, 又可随时读出信息。根据工作原理的不同, RAM 可分为 SRAM 和 DRAM 两类, 两种存储器的存储单元电路分别如图 3.4 和图 3.5 所示。

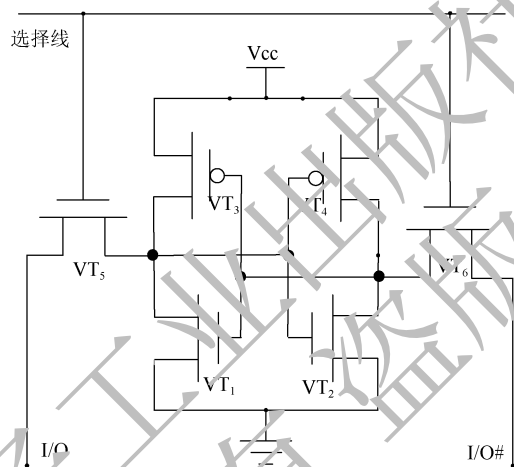


图 3.4 SRAM 的存储单元电路

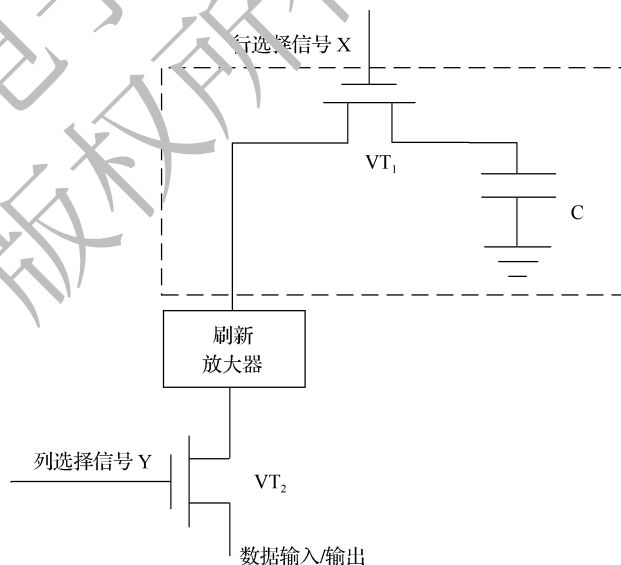


图 3.5 DRAM 的存储单元电路

① SRAM 的存储单元电路。

SRAM 的存储单元电路由 6 个晶体管组成。其中, $VT_1 \sim VT_4$ 组成双稳态锁存器, 用于记忆 1 位二值代码; 控制管 VT_5 、 VT_6 用于实现状态控制。

② DRAM 的存储单元电路。

DRAM 的存储单元电路由 1 个晶体管和 1 个电容组成。数据以电荷的形式存储在栅极电容上, 电容上有电荷代表存储数据 1, 无电荷代表存储数据 0。由于存在漏电, 因此必须定时给电容补充电荷, 以防止信息丢失, 此操作被称为刷新。由于要不断进行刷新, 因此称其为动态存储。

SRAM 的存储单元电路的优点是工作稳定、不需要刷新、运行速度快; 缺点是集成度不高且功耗较大。DRAM 的存储单元电路的优点是集成度高、功耗低, 缺点是运行速度慢, 约为 SRAM 存储单元电路运行速度的一半, 需要刷新。通常缓存使用 SRAM, 内存使用 DRAM。

3.1.2 存储器结构

1. 基本结构

存储器包含三部分: 存储阵列、译码驱动电路和读/写电路。其依靠地址线、数据线和控制总线(片选线和读/写控制线)与外部连接, 基本结构如图 3.6 所示。

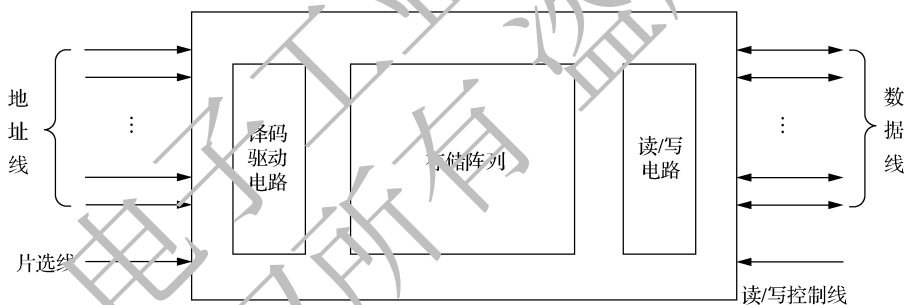


图 3.6 存储器的基本结构

- 片选线: 用于选择芯片。
- 地址线: 单向地址输入。
- 数据线: 双向数据传输。
- 译码驱动电路: 将输入地址翻译成相应存储单元的选择信号。
- 读/写控制线: 决定芯片的读/写操作。
- 读/写电路: 控制读/写操作。
- 存储阵列: 二维矩阵, 按其排列方式可分为两种结构: 多字 1 位结构 (N 字 $\times$ 1 位) 和多字多位结构 (常见的有 N 字 $\times$ 4 位, N 字 $\times$ 8 位)。

(1) 译码驱动方式。

有两种译码驱动方式用于存储单元的选择。



① 线选法。

字选择线（字线）用于选中存储单元（如一字节），位线则是数据线。当存储单元很多时，需要大量字线，所以线选法只适用于存储量不大的芯片。

在图 3.7 中，4 根地址线产生 16 根字线，可以选中 16 个存储单元；8 根位线用于传输每个存储单元的 8 位数据。

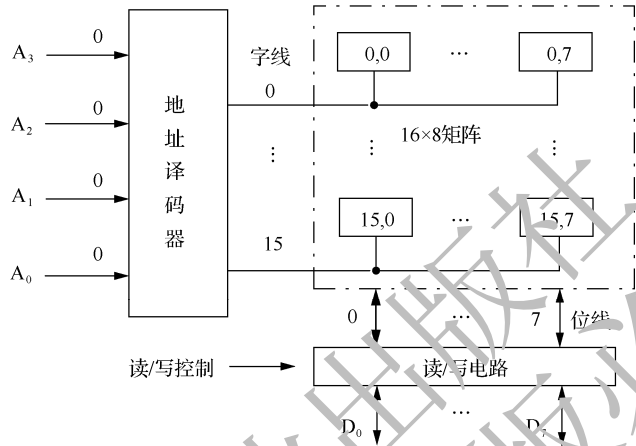


图 3.7 线选法

② 重合法。

在图 3.8 中，地址译码器分成 X（行）地址译码器和 Y（列）地址译码器，在行和列的交汇点处锁定一个存储单元，因此重合法适用于存储量大的芯片。

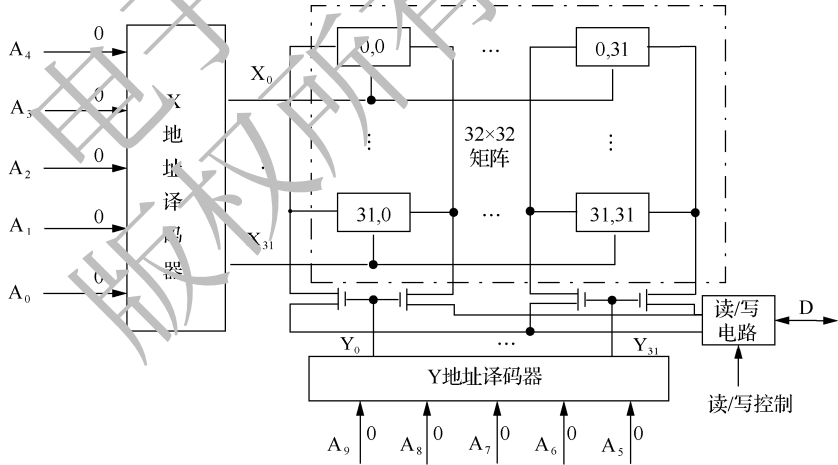


图 3.8 重合法

(2) 寻址原理。

存储器可由多个存储阵列组成。当访问某个存储单元时，根据行选择信号和列选择信号从存储阵列中选中相应的存储单元。

地址线数=行地址线数+列地址线数

存储单元数量=行数×列数

存储容量=存储单元数量×存储单元位宽

假定某存储器共有 24 根地址线，其中 14 根用于行译码，10 根用于列译码，存储单元位宽为 16 位，则可通过计算得到

$$\text{存储单元数量} = 2^{14} \times 2^{10} = 16\text{M}$$

$$\text{存储容量} = 16\text{M} \times 16 = 256\text{Mb} = 32\text{MB}$$

由此可知，该存储器的地址范围为 0x0000_0000（首地址）~0x0FFF_FFFF（末地址）。

2. 存储器扩展

将多个单块 RAM 进行组合，可以扩展成大容量存储器。

（1）位扩展。

位扩展是指在存储器字数不变的前提下进行数据位数的扩展。

将 1MB×1 位的 RAM 扩展为 1MB×8 位的 RAM，并与处理器总线连接，如图 3.9 所示。将 8 颗芯片的数据线连接起来，形成 8 位数据总线 D₇~D₀ 的不同位；各芯片的地址线 A₁₉~A₀ 与处理器地址总线中对应的地址线相连，读/写控制线 R/W 与处理器读/写控制线相连，片选线 CS 与处理器片选线相连。

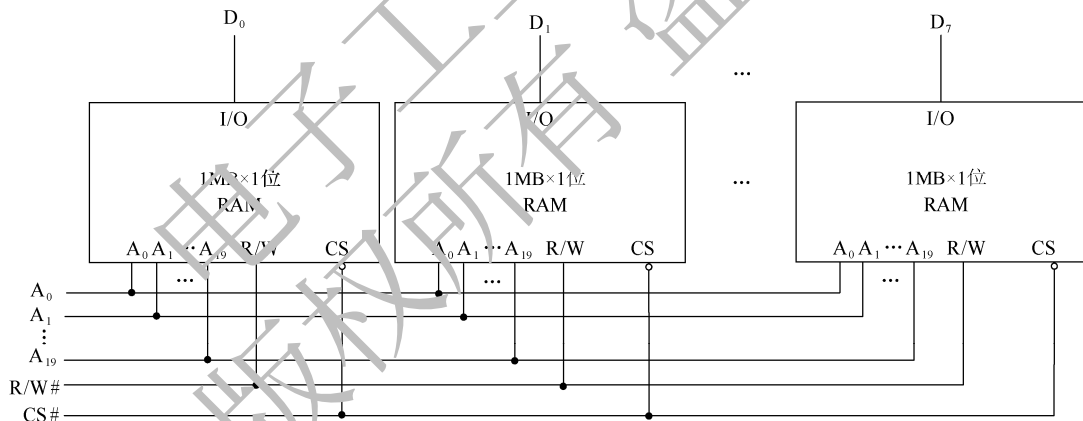


图 3.9 位扩展

（2）字扩展。

字扩展是指在数据位数不变的前提下进行字数扩展，即用扩展的地址线控制各芯片的片选线。

将 256KB×8 位的 RAM 扩展为 1MB×8 位的 RAM，并与处理器总线连接，如图 3.10 所示。各芯片的数据线分别与处理器数据总线 D₇~D₀ 相连；处理器地址总线中的低位地址线 A₁₇~A₀ 与各芯片对应的地址线相连；高位地址线 A₁₉、A₁₈ 通过 2 线-4 线译码器分别产生



不同的译码输出信号，控制各芯片的片选端 CS；各芯片的读/写控制线 R/W 与处理器读/写控制线相连。

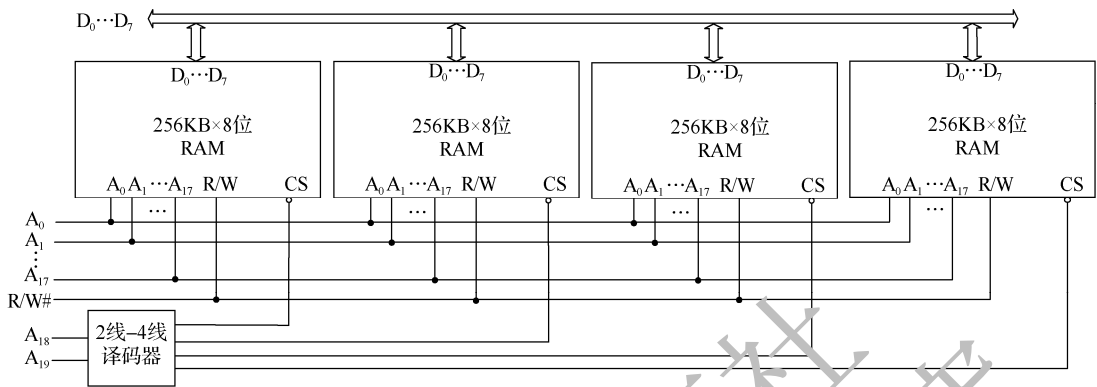


图 3.10 字扩展

(3) 复合扩展。

复合扩展是指对 RAM 的位线和字线都进行扩展，通常先进行位扩展，再进行字扩展。

假定需要将 256KB×1 位的 RAM 扩展为 1MB×8 位的 RAM。先进行位扩展，用 8 个 256KB×1 位 RAM 进行位扩展，构成 256KB×8 位的 RAM，再将位扩展后的 RAM 作为整体进行字扩展，采用 4 个 256KB×8 位的 RAM 构成 1MB×8 位的 RAM。

3. 存储器类型

(1) 单端口 RAM。

单端口 RAM (Single Port RAM) 只有一个读/写端口，即只有一组数据线和地址线端口，读和写都通过此端口来访问，但同一时刻只能进行一种访问，要么是读，要么是写，不能同时读、写。图 3.11 中的单端口 RAM 有 1 个时钟输入端 (CLK)、1 个地址输入端 (ADDR)、1 个写数据输入端 (LIN)、1 个读数据输出端 (DOUT)、1 个使能输入端 (EN) 和 1 个写使能输入端 (WE)。

(2) 伪双端口 RAM。

伪双端口 RAM (Pseudo Dual Port RAM) 有 2 个读/写端口：一个端口只能读；另一个端口只能写。从整体上来看，读、写可以同时进行。伪双端口 RAM 又称为简单双端口 RAM (Simple Dual Port RAM)。图 3.12 所示的伪双端口 RAM 有 2 个时钟输入端 (CLKA、CLKB)、1 个写地址输入端 (ADDRB)、1 个读地址输入端 (ADDRA)、1 个写数据输入端 (DINB)、1 个读数据输出端 (DOUTA)、2 个使能输入端 (ENA、ENB) 和 1 个写使能输入端 (WEB)。

(3) 真双端口 RAM。

真双端口 RAM (True Dual Port RAM) 有 2 个读/写端口，每个端口都可以独立进行读/写操作，既可以同时读或者同时写，又可以一个端口读、一个端口写。图 3.13 所示的真双端

口 RAM 有 2 个时钟输入端 (CLKA、CLKB)、2 个地址输入端 (ADDRA、ADDRB)、2 个写数据输入端 (DINA、DINB)、2 个读数据输出端 (DOUTA、DOUTB)、2 个使能输入端 (ENA、ENB) 和 2 个写使能输入端 (WEA、WEB)。

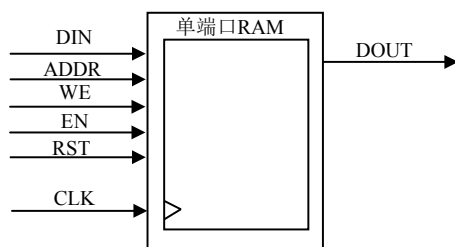


图 3.11 单端口 RAM

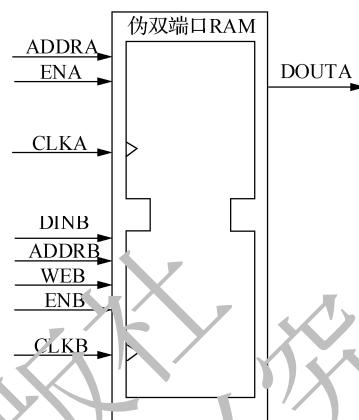


图 3.12 伪双端口 RAM

(4) FIFO 存储器。

FIFO 存储器有 2 个读/写端口，其中一个端口只读，另一个端口只写，如图 3.14 所示。读操作与写操作可以异步进行，数据按照写入顺序被读出。存储容量较大的 FIFO 存储器通常由 RAM 实现，存储容量较小的 FIFO 存储器可以由寄存器实现。

FIFO 存储器与伪双端口 RAM 的区别在于，FIFO 存储器为先入先出，没有地址线，不能对存储单元寻址；伪双端口 RAM 的 2 个读/写端口都有地址线，可以对存储单元寻址。

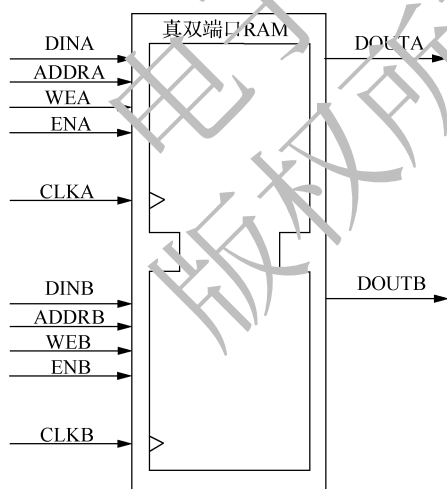


图 3.13 真双端口 RAM

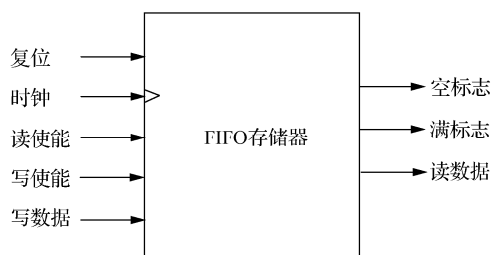


图 3.14 FIFO 存储器

(5) 多端口 RAM。

多端口 RAM (Multi-Port RAM) 可以提供 3 个或更多的读/写端口。