

第5章 存储器系统

【学习目标】

本章首先以半导体存储器为对象,在讨论存储器及其基本电路、基础知识的基础上,讨论存储芯片及其与 CPU 之间的连接和扩充问题;然后,介绍内存的技术发展以及外部存储器;最后,简要介绍存储器系统的分层结构。

【学习要求】

- ◆ 熟悉存储器的分类、组成及功能,着重理解行选与列选对一位信息的读出。
- ◆ 重点掌握位扩充与地址扩充技术。
- ◆ 理解存储器与 CPU 的连接方法。
- ◆ 着重理解内存技术的发展。
- ◆ 理解存储器系统的分层结构概念。

5.1 存储器的分类与组成

计算机的存储器可分为两大类:一类为内部存储器,简称内存或主存,其基本存储元件多以半导体材料制造;另一类为外部存储器,简称外存,多以磁性材料或光学材料制造。

5.1.1 半导体存储器的分类

半导体存储器按使用的功能可分为两大类:随机存取存储器(random access memory, RAM)和只读存储器(read only memory, ROM)。

RAM 按工艺又可分为双极型 RAM 和 MOS RAM 两类,而 MOS RAM 又可分为静态和动态 RAM 两种。

只读存储器(ROM)按工艺也可分为双极型和 MOS 型两类,但一般根据信息写入的方式不同,而分为掩膜式 ROM,可编程 ROM(PROM)和可擦除、可再编程 ROM(紫外线擦除 EPROM 与电子擦除 EEPROM 以及 Flash ROM)等几种。

5.1.2 半导体存储器的组成

半导体存储器一般是由存储体、地址选择电路、读写电路与控制电路组成。

1. 存储体

存储体是存储 1 或 0 信息的电路实体,它由许多个存储单元组成,对每个存储单元赋予一个编号,称为地址单元号。而每个存储单元由若干相同的位组成,每个位需要一个存储元件。

存储器的地址用一组二进制数表示,其地址线的位数 n 与存储单元的数量 N 之间的关系为 $N=2^n$ 。

2. 地址选择电路

地址选择电路包括地址码缓冲器、地址译码器等。

地址译码器用来对地址码译码。设其输入端的地址线根数为 n ,输出线数为 N ,则它分别对应 2^n 个不同的地址码,作为对存储体地址单元的选择线。这些输出的选择线又称字线。

地址译码方式有两种。

1) 单译码方式(或称字结构)

单译码方式的全部地址码只用一个地址译码器电路译码,译码输出的字选择线直接选中与输入地址码对应的存储单元。单译码方式需要的选择线数较多,只适用于容量较小的存储器。

2) 双译码方式(或称重合译码)

双译码方式是将地址码分为 X 与 Y 两部分,用两个译码电路分别译码。 X 向译码又称行译码,其输出线称为行选择线,它选中存储矩阵中一行的所有存储单元。 Y 向译码又称列译码,其输出线称为列选择线,它选中存储矩阵中一列的所有存储单元。只有 X 向和 Y 向的选择线同时选中的那一位存储单元,才能进行读或写操作。由于双译码方式所需要的选择线数目较少,也简化了存储器的结构,故它适用于大容量的存储器。

3. 读写电路与控制电路

读写电路包括读写放大器、数据缓冲器(三态双向缓冲器)等,是数据信息输入和输出的通道。

外界对存储器的控制信号有读信号($\overline{RD}$)、写信号($\overline{WR}$)和片选信号($\overline{CS}$)等,通过控制电路以控制存储器的读或写操作以及片选。只有片选信号处于有效状态时,存储器才能与外界交换信息。

5.2 随机存取存储器

RAM 既可以读出,也可以写入。RAM 可分为静态内存(static RAM, SRAM)和动态内存(dynamic RAM, DRAM)两大类。常用静态内存作为系统的高速缓存(通常用于

一级缓存和二级缓存),而平常所提到的内存指的是动态内存。

5.2.1 静态随机存取存储器

1. SRAM 基本存储电路

SRAM 的基本存储电路,是由 6 个 MOS 管组成的 RS 触发器。每一个触发器就构成存储体的一位。

2. SRAM 的组成

SRAM 一般由存储体、译码电路和控制电路组成。

存储体由存储矩阵构成。在存储矩阵中,只有行、列均被选中的某个单元存储电路(即一位),在其 X 向选通门与 Y 向选通门同时被打开时,才能进行读出信息和写入信息的操作。

存储体中的每一位,仅有一个 I/O 电路用于存取各存储单元中的一位信息。如果要组成字长为 4 位或 8 位的存储器,则 I/O 电路也应有 4 个或 8 个。这样,当存储体的某个存储单元在一次存取操作中被地址译码器输出端的有效输出电平选中时,则该单元内的 4 位或 8 位信息将被一次读写完毕。

通常,一个 RAM 芯片的存储容量是有限的,需要用若干片才能构成一个实用的存储器。这样,地址不同的存储单元,可能处于不同的芯片中,因此,在选中地址时,应先选择其所属的芯片。对于每块芯片,都有一个片选控制端($\overline{CS}$),只有当片选端加上有效信号时,才能对该芯片进行读或写操作。一般片选信号由地址码的高位译码(通过译码器输出端)产生。

3. SRAM 的读写过程

SRAM 的读写过程参见主教材的图 5-5。

1) 读出过程

(1) 地址码 $A_0 \sim A_{11}$ 加到 RAM 芯片的地址输入端,经 X 与 Y 地址译码器译码,产生行选与列选信号,选中某一存储单元,该单元中存储的代码,经一定时间,出现在 I/O 电路的输入端。I/O 电路对读出的信号进行放大、整形,送至输出缓冲寄存器。缓冲寄存器一般具有三态控制功能,没有开门控制信号,所存数据还不能送到数据总线 DB 上。

(2) 在送上地址码的同时,还要送上读写控制信号($R/\overline{W}$ 或 $\overline{RD}$ 、 $\overline{WR}$)和片选信号($\overline{CS}$)。读出时,使 $R/\overline{W}=1$, $\overline{CS}=0$,这时,输出缓冲寄存器的三态门将被打开,所存信息送至 DB 上。于是,存储单元中的信息被读出。

2) 写入过程

(1) 同上述读出过程(1),先选中相应的存储单元,使其可以进行写操作。

(2) 将要写入的数据放在 DB 上。

(3) 加上片选信号 $\overline{CS}=0$ 及写入信号 $R/\overline{W}=0$ 。这两个有效控制信号打开三态门,使 DB 上的数据进入输入电路,送到存储单元的位线上,从而写入该存储单元。

需要着重理解,无论是存储器读还是存储器写,都必须保证能从存储器中读出或写入

存储器中一个稳定的数据。在读写信号有效期间,数据线和地址线也必须是稳定的。

4. SRAM 芯片举例

常用的 SRAM 芯片有 2114、2142、6116、6264 等。

本书选用了 Intel 6116 和 6264。6116 是一个 $2K \times 8$ 位的 CMOS SRAM 芯片,属双列直插式、24 条引脚封装。它的存储容量为 $2K \times 8$ 位。6264 芯片的结构及工作原理与 6116 芯片相似,是一个存储容量为 $8K \times 8$ 位的 CMOS SRAM 芯片,其引脚有 28 条。

5.2.2 动态随机存取存储器

动态随机存取存储器(DRAM)芯片是以 MOS 管栅极电容是否充有电荷来存储信息的,其基本单元电路一般由四管、三管和单管组成,以三管和单管较为常用。由于它所需要的管子较少,故可以扩大每片存储器芯片的容量,并且其功耗较低,所以在微型计算机系统中,大多数采用 DRAM 芯片。

1. 动态基本存储电路

这里重点介绍常用的三管和单管这两种基本存储电路。

1) 三管动态基本存储电路

三管动态基本存储电路由 T_1 、 T_2 、 T_3 三个管子和两条字选择线(读、写选择线)、两条数据线(读、写数据线)组成。 T_1 是写数控制管; T_2 是存储管,用它的栅极电容 C_g 存储信息; T_3 是读数控制管; T_4 管是一列基本存储电路上共同的预充电管,以控制对输出电容 C_D 的预充电。

写入操作时,信息被保存在电容 C_g 上。

读出操作时,在读数据线上可以读出与原存储相反的信息。若再经过读出放大器反相后,就可以得到原存储信息了。

对于三管动态基本存储电路,必须每隔 $1 \sim 3ms$ 定时对 C_g 充电,以保持原存信息不变,此即动态存储器的刷新(或称为再生)。

刷新要由刷新电路来实现。

2) 单管动态基本存储电路

单管动态基本存储电路仅由一个 T_1 管和寄生电容 C_s 组成。

写入时,使字选线上为高电平, T_1 管导通,待写入的信息由位线 D(数据线)存入 C_s ;读出时,存储在 C_s 上的信息通过 T_1 管送到 D 线上,再通过放大,即可得到存储信息。由于单管动态基本存储电路使用管子少,4KB 以上容量较大的 RAM,大多采用单管电路。

2. 动态 RAM 芯片举例

Intel 2116 是单管动态 RAM 芯片。它的存储容量为 $16K \times 1$ 位。本来这需用 14 条地址输入线,但由于受封装引线的限制,2116 芯片只有 16 条引脚,其中只有 7 条地址输入线。

为了解决用 7 条地址输入线传送 14 位地址码的矛盾,2116 芯片采用地址线分时复用技术。

2116 芯片没有片选信号 $\overline{CS}$,它的行地址选通信号 $\overline{RAS}$ 兼作片选信号,且在整个读、写周期中均处于有效状态,这是与其他芯片的不同之处。

此外,地址输入线 $A_0 \sim A_6$ 还用作刷新地址的输入端,刷新地址由 CPU 内部的刷新寄存器 R 提供。

与 Intel 2116 芯片类似的还有 2164、3764、4164 等 DRAM 芯片。

5.3 只读存储器

5.3.1 只读存储器存储信息的原理和组成

ROM 的存储元件可以看作一个单向导通的开关电路。

ROM 的组成结构与 RAM 相似,一般也是由地址译码电路、存储矩阵、读出电路及控制电路等部分组成的。

5.3.2 只读存储器的分类

1. 不可编程掩膜式 MOS 只读存储器

不可编程掩膜式 MOS ROM 又称固定存储器。它是由器件制造厂家根据用户编好的机器码程序,把 0、1 信息存储在掩膜图形中而制成的 ROM 芯片。这种芯片制成以后,它的存储矩阵中每个 MOS 管所存储的信息 0 或 1 被固定下来,不能再改变,而只能读出。如果要修改其内容,那么只有重新制作。

2. 可编程只读存储器

为了克服上述掩膜式 MOS ROM 芯片不能修改内容的缺点,设计了一种可编程序的只读存储器(programmable ROM, PROM),用户在使用前可以根据自己的需要编制 ROM 中的程序。

3. 可擦除、可再编程的只读存储器

PROM 芯片虽然可供用户进行一次修改程序,但有很大局限。为了便于研究工作,试验各种 ROM 程序方案,就研制了一种可擦除、可再编程的 ROM(erasable PROM, EPROM)。在 EPROM 芯片出厂时,它是未编程的。若 EPROM 中写入的信息有错或不需时,可利用专用的紫外线灯对准芯片上的石英窗口照射 15~20min,即可擦除原写入的信息。

还有一种方法是采用金属-氮-氧化物-硅(MNOS)工艺生产的 MNOS 型 PROM,它是一种利用电来改写的可编程只读存储器,即 EEPROM(或称 E^2 PROM)。用这种方法可改写数万次,且每次只需要 0.1~0.6s,信息存储时间可达十余年之久,这给需要经常修改程序和参数的应用领域带来极大的方便。但是, E^2 PROM 有存取时间较慢,完成改写程序需要较复杂的设备等缺点,现在正在迅速发展和应用高密度、高存取速度的

E²PROM 技术和闪存(flash memory)技术。

5.3.3 常用 ROM 芯片举例

1. Intel 2732 芯片

2732 EPROM 芯片的容量为 $4\text{K} \times 8$ 位,采用 HN MOS-E(高速 NMOS 硅栅)工艺制造和双列直插式封装,其引脚有 24 条。

与 2732 芯片属于同一类的常用 EPROM 芯片还有 2764、27128、27256、27512、271024 等,它们的内部结构与外部引脚分配基本相同,主要是存储容量逐次成倍递增为 $4\text{K} \times 8$ 位、 $8\text{K} \times 8$ 位、 $16\text{K} \times 8$ 、 $32\text{K} \times 8$ 位、 $64\text{K} \times 8$ 位、 $128\text{K} \times 8$ 位等。

2. E²PROM 芯片举例

常用的 E²PROM 芯片有 2816/2816A、2817/2817A、2864A 等。其中,以 2864A 的 $8\text{K} \times 8$ 位容量大,且与 6264 兼容。其主要特点是能像 SRAM 芯片一样读写操作,读访问时间可为 $45 \sim 450\text{ns}$,在写之前自动擦除原内容。但它并不能像 RAM 芯片那样随机读写,而只能有条件地写入,即只有当一字节或一页数据编程写入结束后,方可以写入下一字节或下一页数据。在 E²PROM 的应用中,若需读其某个单元的内容,只要执行一条存储器读指令,即可读出;若需对其内容重新编程,可在线直接用字节写入或页写入方式写入。

3. Flash ROM 芯片

在 Pentium CPU 以上的主板中普遍采用了 Flash ROM 芯片来作为 BIOS 程序的载体。Flash ROM 也称为闪速存储器,在本质上属于 E²PROM。

5.4 存储器的扩充及其与 CPU 的连接

存储器的连接主要解决两个问题:一个是如何用容量较小、字长较短的芯片,组成微机系统所需的存储器;另一个是存储器与 CPU 的连接方法与应注意的问题。

5.4.1 存储器芯片的扩充技术

1. 位扩充

一块实际的存储芯片,其存储单元的位数(即字长)通常与实际内存单元的字长并不相等,如 SRAM 芯片 2114 的 $1\text{K} \times 4$ 位,DRAM 芯片 2164 为 $64\text{K} \times 1$ 位等。显然,要用这些芯片构成实际上按字节组织的内存空间,就需要进行位的扩充,以满足字长的要求。

用 1 位或 4 位的存储器芯片构成 8 位字长的存储器,可采用位并联的方法。例如,可以用两片 $4\text{K} \times 4$ 位的存储器芯片经过位扩充构成 4KB 的存储器。

2. 字扩充

地址的扩充即存储容量的扩充(又称字扩充)。当扩充存储容量时,采用地址串联的方法。这时,要用到地址译码电路,以其输入的地址码来区分高位地址,而以其输出端的

控制线来对具有相同低位地址的几片存储器芯片进行片选。

地址译码电路是一种可以将地址码翻译成相应控制信号的电路,有 2-4 译码器、3-8 译码器等。

当需要同时位扩充与字扩充时,可以将上述两种方法结合起来使用。例如,当用 $16\text{K} \times 1$ 位的芯片组成 $64\text{K} \times 8$ 位的存储器时,共需用 32 片 $16\text{K} \times 1$ 位的芯片。先用位线并联方法将每 8 片组成一组 $16\text{K} \times 8$ 位存储器,再用字扩充方法,选用 2-4 译码器组成的译码电路,组成 4 组 $16\text{K} \times 8$ 位共 $64\text{K} \times 8$ 位存储器。

5.4.2 存储器与 CPU 的连接

1. 只读存储器与 8086 CPU 的连接

ROM、PROM、EPROM 芯片都可以与 8086 系统总线连接,以组成程序存储器。例如,Intel 2716、2732、2764、27128 这一类 EPROM 芯片,由于它们是属于以 1 字节宽度输出组织的,因此,在连接到 8086 系统时,为了存储 16 位指令字,要使用两片这类芯片并联组成一组。

注意:在由两片 2732 组成的 4K 字程序存储器中,两片的片选端 $\overline{\text{CE}}$ 一起连到译码器的片选端 $\overline{\text{CS}}$ 。

2. 静态 RAM 与 8086 CPU 的连接

一般,当微型计算机系统的存储器容量少于 16K 字时,宜采用静态 RAM 芯片,因为大多数动态 RAM 芯片都是以 $16\text{K} \times 1$ 位或 $64\text{K} \times 1$ 位来组织的,并且动态 RAM 芯片还要求动态刷新电路,这种附加的支持电路会增加存储器的成本。

8086 CPU 无论是在最小模式或最大模式下,都可以寻址 1MB 的存储单元,存储器均按字节编址。

注意:在由两片 6116 组成的 2K 字数据存储器中,两片的片选端 $\overline{\text{CE}}$ 分别连到系统的 A_0 、 $\overline{\text{BHE}}$ 端。

3. EPROM、SRAM 与 8086 CPU 连接的实例

主教材中的图 5.24 给出了 8086 单处理器系统连接实例,它就是一个实际单板机的电路图。深入理解本系统的原理和电路,就能综合掌握和灵活运用有关存储器与 CPU 的连接知识。

4. 32 位或 64 位存储器接口

32 位存储器接口或 64 位存储器接口原理同上面介绍的 16 位存储器接口基本一样。其主要区别在于 32 位存储器接口与 64 位存储器接口需要的存储体个数分别为 4 个与 8 个。此外,由于 16 位、32 位或 64 位微处理器地址线数目的不同,分别所能寻址的空间大小是不同的。例如,在 32 位存储器接口中,微处理器有 32 条地址线,其寻址空间为 4GB;而在 Pentium 系统中,微处理器可以被设置为 36 条地址线,其最大寻址空间则为 64GB。至于在存储器接口中,存储体与微处理器之间的具体连接方法还涉及译码器的选用与连接等问题。

5.5 内存的技术发展

内存历来都是系统中最大的性能瓶颈之一。随着 PC 技术的发展,内存条从规格、技术到总线带宽等不断更新换代,使内存的性能瓶颈问题获得较大改善。图 5-1 为多个内存条的针脚与接口设计的示意图。

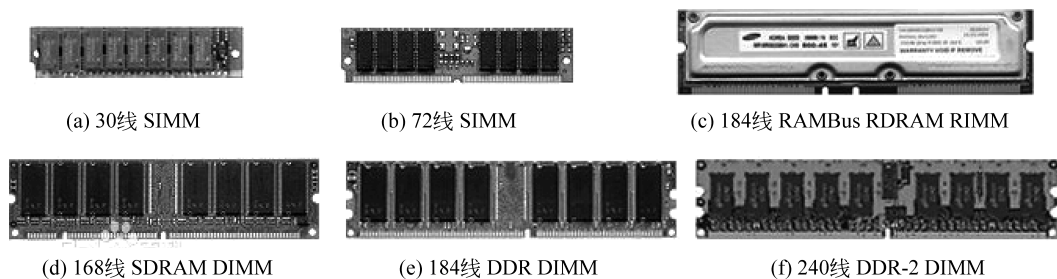


图 5-1 内存条的针脚与接口设计

1. SIMM 内存

1982 年最初出现在 80286 主板上的“内存条”,采用的是单边接触内存模组(single inline memory modules, SIMM)接口,容量为 30 线、256KB,由 8 片数据位和 1 片校验位组成 1 个存储区块(bank),因此,一般见到的 30 线 SIMM 都是 4 条一起使用的。

在 1988 年至 1990 年,PC 技术进入 32 位的 386 和 486 时代,推出了 72 线 SIMM 内存,它支持 32 位快速页模式内存,内存带宽得以大幅度提升。72 线 SIMM 内存单条容量一般为 512KB ~ 2MB,要求两条同时使用。

2. EDO DRAM 内存

外扩充数据模式动态存储器(extended data out DRAM, EDO DRAM)是 1991 年至 1995 年盛行的内存条。工作电压一般为 5V,带宽 32 位,主要应用在当时的 486 及早期的 Pentium 计算机中。随着 EDO DRAM 在成本和容量上的突破,加上制作工艺的飞速发展,当时单条 EDO DRAM 内存的容量已达到 4~16MB。

3. SDRAM 时代

自 Intel Celeron 系列和 AMD K6 处理器以及相关的主板芯片组推出后,EDO DRAM 内存性能再也无法满足需要,于是内存又开始进入比较经典的 SDRAM 时代。

第一代 SDRAM 内存为 PC66 规范,之后有 PC100、PC133、PC150 等规范,其频率从早期的 66MHz,发展到 100MHz、133MHz 等。由于 SDRAM 的带宽为 64 位,正好对应 CPU 的 64 位数据总线宽度,因此,它只需要一条内存便可工作,便捷性进一步提高。

4. Rambus DRAM 内存

Intel 公司在推出高频 Pentium III 和 Pentium 4 CPU 的同时,推出了 Rambus DRAM 内存。Rambus DRAM 内存以高时钟频率来简化每个时钟周期的数据量,因此,

内存带宽相当出色,如 PC 1066,1066MHz、32 位带宽可达到 4.2GB/s,它曾一度被认为是 Pentium 4 的绝配。

5. DDR 时代

双倍速率 SDRAM(double data rate SDRAM,DDR SDRAM,简称 DDR),实际上是 SDRAM 的升级版,在时钟信号的上升沿和下降沿都可以传输数据,因而时钟率可以加倍提高,传输速率和带宽也相应提高。

内存发展到 SDRAM 末期,出现了 RDRAM 和 DDR 的路线之争。RDRAM 的技术核心是串行,DDR 的技术核心是数据预取。

DDR SDRAM 内存有 184 条引脚,第一代 DDR200 规范未得到普及,第二代 PC266 DDR SDRAM(133MHz 时钟 $\times 2$ 倍数据传输=266MHz 带宽)是由 PC133 SDRAM 内存衍生而来(不少赛扬和 AMD K7 处理器都采用了 DDR266 规格的内存),其后来的 DDR333 内存也属于一种过渡;双通道 DDR400 内存已经成为前端总线 800FSB 处理器搭配的基本标准。

6. DDR2 时代

随着 CPU 性能的不断提高,对内存性能的要求也逐步升级。针对 PC 等市场的 DDR2 内存拥有 400MHz、533MHz、667MHz 等不同的时钟频率,高端的 DDR2 内存速度已经提升到 800MHz/1066MHz。DDR2 内存采用 200/220/240 针脚的 FBGA 封装形式,它可以提供更良好的电气性能与散热性。LGA775 接口的 915/925 和 945 等平台都支持 DDR2 内存。

7. DDR3 时代

2007 年 JEDEC 确定了 DDR3 内存规范。DDR3 在 DDR2 基础上采用新型设计,其工作电压更低,从 DDR2 的 1.8V 降落到 1.5V,性能更好、更省电;从 DDR2 的 4 位预读升级为 8 位预读;等效频率从 DDR3 800 提升到 DDR3 1600 甚至 DDR3 2133。

面向 64 位构架的 DDR3 显然在频率和速度上拥有更多的优势。市场对 DDR3 内存的需求顶点在 2012 年达成,其市场占有率约为 71%。

8. DDR4 时代

2012 年 JEDEC 又发布了新的 DDR4 规范。JEDEC 的内存规范极其详尽,包括芯片设计、PCB 层数、频率等重要参数。

内存发展的核心目标是不断提升速度。DDR 在发展过程中,一直都以增加数据预取值作为主要的性能提升手段。在 DDR3 上使用了 8 位预取。预取在已达到 $8n$ 的情况下难以进一步提升。设计者想出了新的办法:在内部设计了 Bank Group 架构。这样的技术只是在内存内部的结构做出优化设计,并没有直接提升最根本的数据预取值,相当于组建了内存内部的多通道。Bank Group 带来了 DDR4 内部数据传输能力的大幅度提升,让 DDR4 在物理频率没有太大提升的情况下大幅度提升数据存取能力。

Bank Group 是 DDR4 提升内存带宽的关键技术,而点对点总线则是 DDR4 整个存储系统的关键性设计。在传统的 DDR3 等内存上,内存和内存控制器连接依靠的是多点分支总线(multi-drop bus)。这种总线允许在一个接口上挂接许多同样规格的芯片。

点对点的优点很多,例如设计比较简单、容易达到更高的频率等。它的问题也很明显:一个重要因素是点对点总线每个通道只能支持一根内存,因此,如果 DDR4 内存单条容量不足,将很难有效提升系统的内存总量。

DDR4 启用了 3DS 堆叠封装技术来增大单个芯片的容量,这也是 DDR4 内存中最关键的技术之一。常见的大容量内存单条容量为 8GB(单个芯片 512MB,共 16 个),而 DDR4 最大容量可以达到 64GB,甚至 128GB,彻底解决了点对点总线容量不足的问题。

总之,DDR4 采用了 Bank Group 技术,再加上全新的点对点传输、大量的功耗控制和信号完整性控制技术,呈现出极为优异的发展态势。图 5-2 为内存的工作电压规格路线图,图 5-3 为内存频率规格的发展示意图。

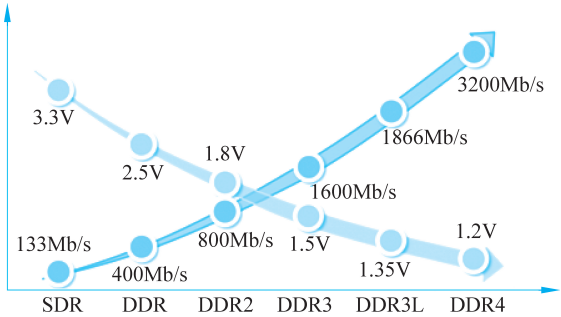


图 5-2 内存的工作电压规格路线图

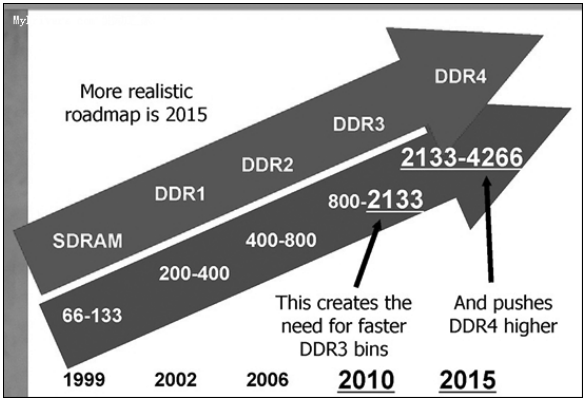


图 5-3 内存频率规格的发展示意图

DDR4 2014 年开始生产,已经取代 DDR3 成为主流产品。

5.6 外部存储器

硬盘(hard disk)是计算机最重要的外部存储设备,包括操作系统在内的各种软件、程序、数据都需要保存在硬盘上,其性能直接影响计算机的整体性能。光盘存储技术是采用磁盘以来最重要的新型数据存储技术,它具有容量大、速度快、工作稳定可靠以及耐用性