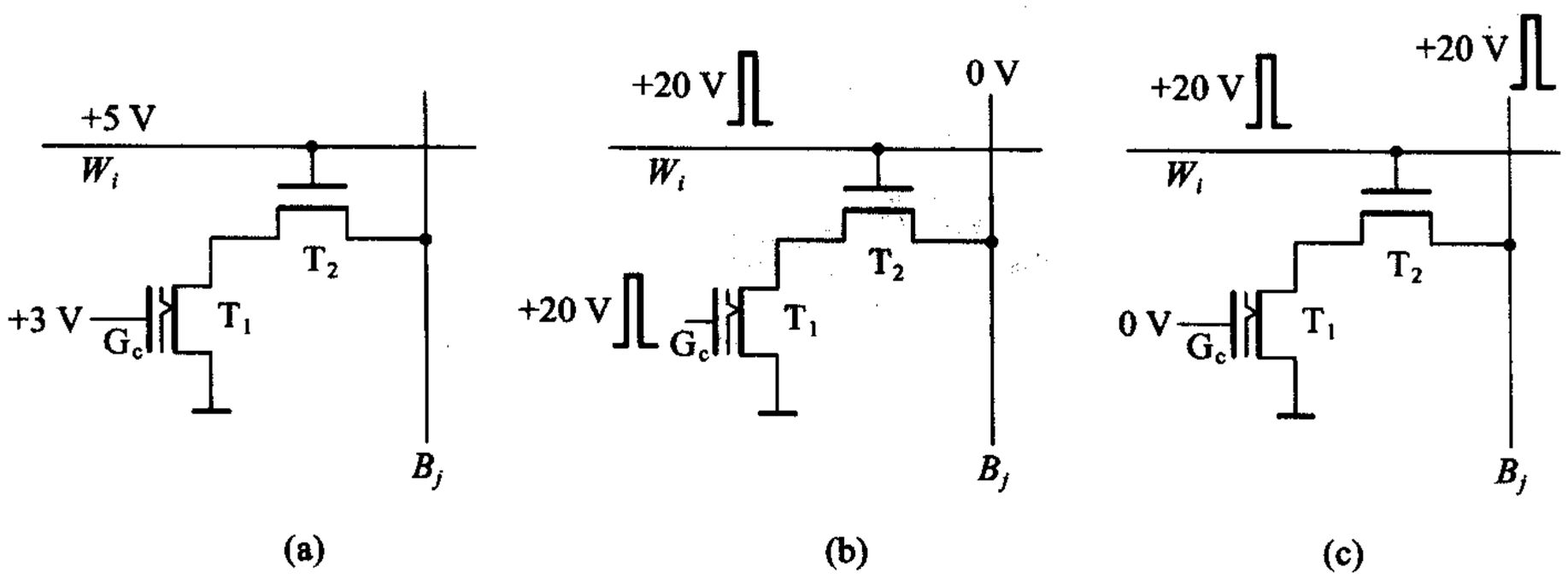


图 7.2.10　E^2PROM存储单元的三种工作状态

(a) 读出状态　(b) 擦除(写 1)状态　(c) 写入(写 0)状态

10 ms 的脉冲电压,漏区接 0 电平,如图 7.2.10(b)所示。这时经 $G_c - G_f$ 间电容和 G_f – 漏区电容分压在隧道区产生强电场,吸引漏区的电子通过隧道区到达浮置栅,形成存储电荷,使Flotox管的开启电压提高到 +7 V 以上,成为高开启电压管。读出时 G_c 上的电压只有 +3 V,Flotox管不会导通。一个字节擦除以后,所有的存储单元均为 1 状态。

在写入状态下,应使写入 0 的那些存储单元的Flotox管浮置栅放电。为此,在写入 0 时令控制栅 G_c 为 0 电平,同时在字线 W_i 和位线 B_j 上加 +20V 左右、宽度约 10 ms 的脉冲电压,如图 7.2.10(c)所示。这时浮置栅上的存储电荷将通过隧道区放电,使Flotox管的开启电压降为 0 V 左右,成为低开启电压管。读出时 G_c 上加 +3 V 电压,Flotox管为导通状态。

虽然E^2PROM改用电压信号擦除了,但由于擦除和写入时需要加高电压脉冲,而且擦、写的时间仍较长,所以在系统的正常工作状态下,E^2PROM仍然只能工作在它的读出状态,作 ROM 使用。

三、快闪存储器(Flash Memory)

从上面对E^2PROM的介绍中可以看到,为了提高擦除和写入的可靠性,E^2PROM的存储单元用了两只MOS管。这无疑将限制了E^2PROM集成度的进一步提高。而快闪存储器则采用了一种类似于EPROM的单管叠栅结构的存储单元,制成了新一代用电信号擦除的可编程 ROM。

快闪存储器既吸收了EPROM结构简单、编程可靠的优点,又保留了E^2PROM用隧道效应擦除的快捷特性,而且集成度可以做得很高。图 7.2.11 是快闪存储器采用的叠栅MOS管的结构示意图。它的结构与EPROM中的 SIMOS 管极为相似,两者最大的区别是浮置栅与衬底间氧化层的厚度不同。在EPROM中这个氧

化层的厚度一般为 30 ~ 40 nm,而在快闪存储器中仅为 10 ~ 15 nm。而且浮栅与源区重叠的部分是由源区的横向扩散形成的,面积极小,因而浮置栅 - 源区间的电容要比浮置栅 - 控制栅间的电容小得多。当控制栅和源极间加上电压时,大部分电压都将降在浮置栅与源极之间的电容上。快闪存储器的存储单元就是用这样一只单管组成的,如图 7.2.12 所示。

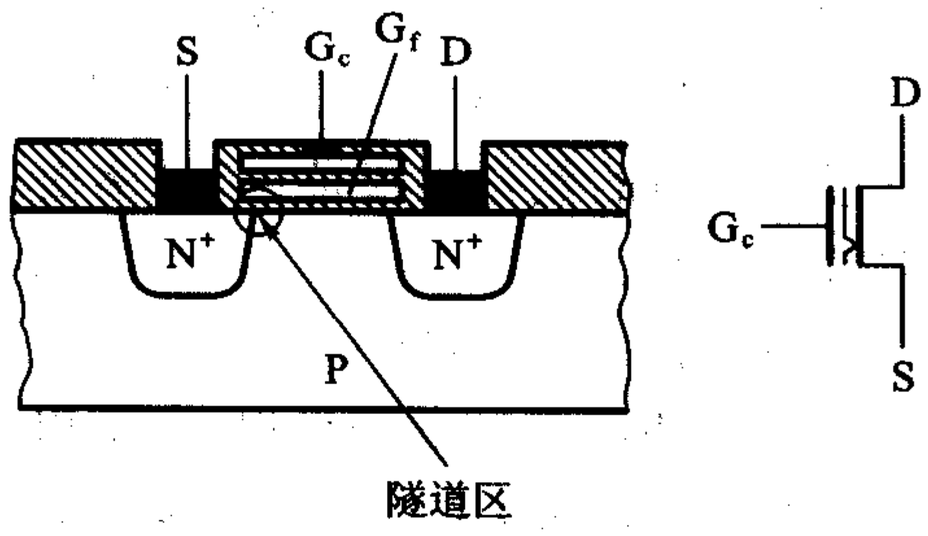


图 7.2.11 快闪存储器中的叠栅MOS管

在读出状态下,字线给出 +5 V 的逻辑高电平,存储单元公共端 V_{SS} 为 **0** 电平。如果浮置栅上没有充电,则叠栅MOS管导通,位线上输出低电平;如果浮置栅上充有负电荷,则叠栅MOS管截止,位线上输出高电平。

快闪存储器的写入方法是利用雪崩注入的方法使浮栅充电的。在写入状态下,叠栅MOS管的漏极经位线接至一个较高的正电压(一般为 6 V),V_{SS} 接 **0** 电平,同时在控制栅上加一个幅度 12 V 左右、宽度约 10 μs 的正脉冲。这时 D - S 间将发生雪崩击穿,一部分速度高的电子便穿过氧化层到达浮置栅,形成浮置栅充电电荷。浮置栅充电后,叠栅MOS管的开启电压为 7 V 以上,字线为正常的逻辑高电平时它不会导通。

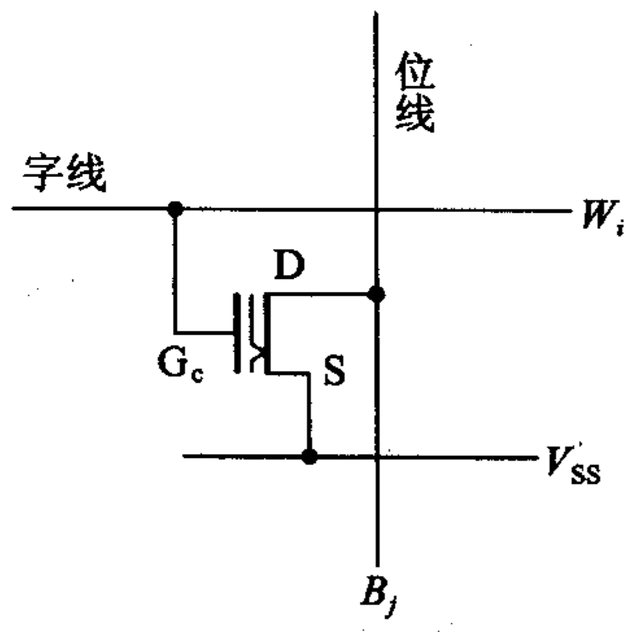


图 7.2.12 快闪存储器的存储单元

快闪存储器的擦除操作是利用隧道效应进行的,在这一点上又类似于 E^2PROM写入 **0** 时的操作。在擦除状态下,令控制栅处于 **0** 电平,同时在源极 V_{SS} 加入幅度为 12 V 左右、宽度约 100 ms 的正脉冲。这时在浮置栅与源区间极小的重叠部分产生隧道效应,使浮置栅上的电荷经隧道区释放。浮置栅放电后,叠栅MOS管的开启电压在 2 V 以下,在它的控制栅上加 +5 V 的电压时一定会导通。

对于采用如图 7.2.3 所示的或门结构形式快闪存储器,其中所有的源极都是连在一起的,所以当在源极上加高电压进行擦除时,所有字线为零的那些字节

中的存储单元将同时被擦除。利用这个特点,在对一个大容量快闪存储器进行擦除时,就可以将它划分成几个区。每次将一个区内的全部存储单元一次同时擦除,从而大大提高了擦除速度。相比之下,编程写入的速度要慢得多。因此,它不能满足随时进行快速写入和读出的要求,通常情况下仍然作为只读存储器使用。

快闪存储器的编程和擦除操作不需要使用编程器,写入和擦除的控制电路集成于存储器芯片中,工作时只需要 5 V 的低压电源,使用极其方便。

由于叠栅MOS管浮置栅下面的氧化层极薄,经过多次编程以后可能发生损坏,所以目前快闪存储器的编程次数是有限的,一般在 10 000 ~ 100 000 次之间。随着制造工艺的改进,可编程的次数有望进一步增加。

自从 20 世纪 80 年代末期快闪存储器问世以来,便以其高集成度、大容量、低成本和使用方便等优点而引起普遍关注。产品的集成度在逐年提高,64 M 位以上的产品已经面市。应用领域迅速扩展,它不仅取代了从前普遍使用的软磁盘,而且有可能在不久的将来成为较大容量磁性存储器(例如 PC 机中的硬磁盘)的替代产品。

复习思考题

R7.2.1 既然快闪存储器能够擦除后改写(重新编程),为什么还把它归类到只读存储器当中呢?

7.3 随机存储器(RAM)

随机存储器也称随机读/写存储器,简称 RAM。在 RAM 工作时可以随时从任何一个指定地址读出数据,也可以随时将数据写入任何一个指定的存储单元中去。它的最大优点是读、写方便,使用灵活。但是,它也存在数据易失性的缺点(即一旦停电以后所存储的数据将随之丢失)。RAM 又分为静态随机存储器 SRAM 和动态随机存储器 DRAM 两大类。

7.3.1 静态随机存储器(SRAM)

一、SRAM 的结构和工作原理

SRAM 电路通常由存储矩阵、地址译码器和读/写控制电路(也称输入/输出

电路)三部分组成,如图 7.3.1 所示。

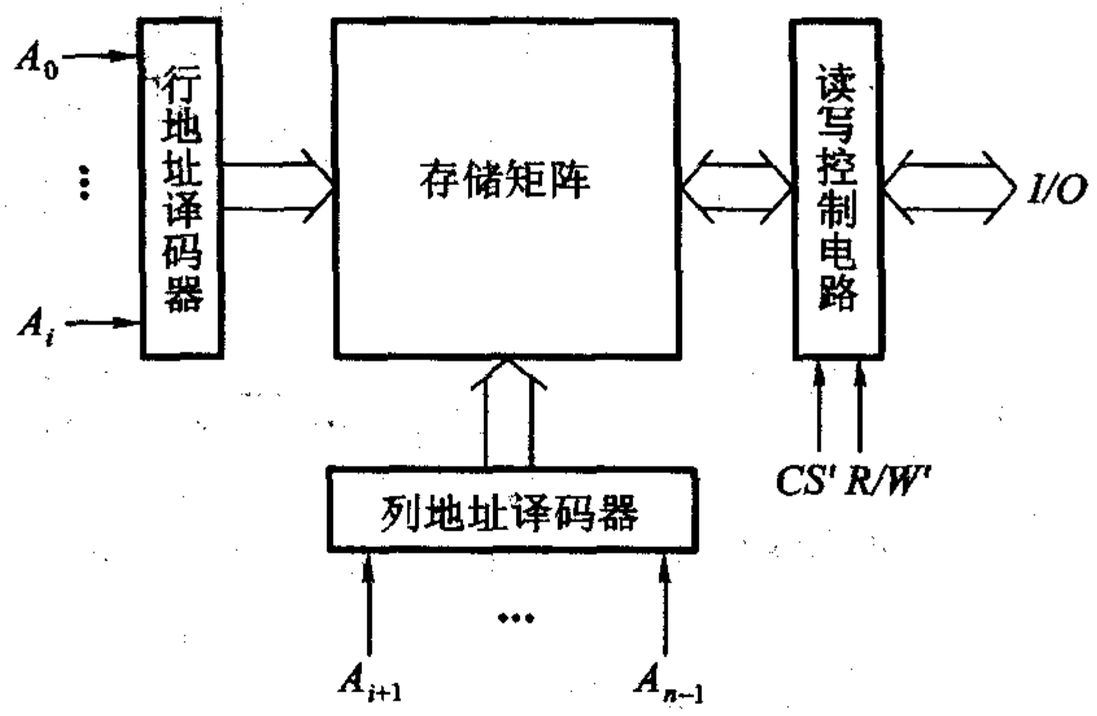


图 7.3.1 SRAM 的结构框图

存储矩阵由许多存储单元排列而成,每个存储单元能存储 1 位二值数据(**1** 或 **0**),在译码器和读/写电路的控制下,既可以写入 **1** 或 **0**,又可以将存储的数据读出。

地址译码器一般都分成行地址译码器和列地址译码器两部分。行地址译码器将输入地址代码的若干位译成某一条字线的输出高、低电平信号,从存储矩阵中选中一行存储单元;列地址译码器将输入地址代码的其余几位译成某一根输出线上的高、低电平信号,从字线选中的一行存储单元中再选 1 位(或几位),使这些被选中的单元经读/写控制电路与输入/输出端接通,以便对这些单元进行读、写操作。

读/写控制电路用于对电路的工作状态进行控制。当读/写控制信号 $R/W'=1$ 时,执行读操作,将存储单元里的数据送到输入/输出端上。当 $R/W'=0$ 时,执行写操作,加到输入/输出端上的数据被写入存储单元中。图中的双向箭头表示一组可双向传输数据的导线,它所包含的导线数目等于并行输入/输出数据的位数。多数 RAM 集成电路是用一根读/写控制线控制读/写操作的,但也有少数的 RAM 集成电路是用两个输入端分别进行读和写控制的。

在读/写控制电路上都设有片选输入端 CS'。当 $CS'=0$ 时 RAM 为正常工作状态;当 $CS'=1$ 时所有的输入/输出端均为高阻态,不能对 RAM 进行读/写操作。

图 7.3.2 是一个 1024 × 4 位 RAM 的实例——2114 的结构框图,其中 4096 个存储单元排列成 64 行 × 64 列的矩阵。10 位输入地址代码分成两组译码。$A_3 \sim A_8$6 位地址码加到行地址译码器上,用它的输出信号从 64 行存储单元中选出指定的一行。另外 4 位地址码加到列地址译码器上,利用它的输出信号再从已选中的一行里挑出要进行读/写的 4 个存储单元。

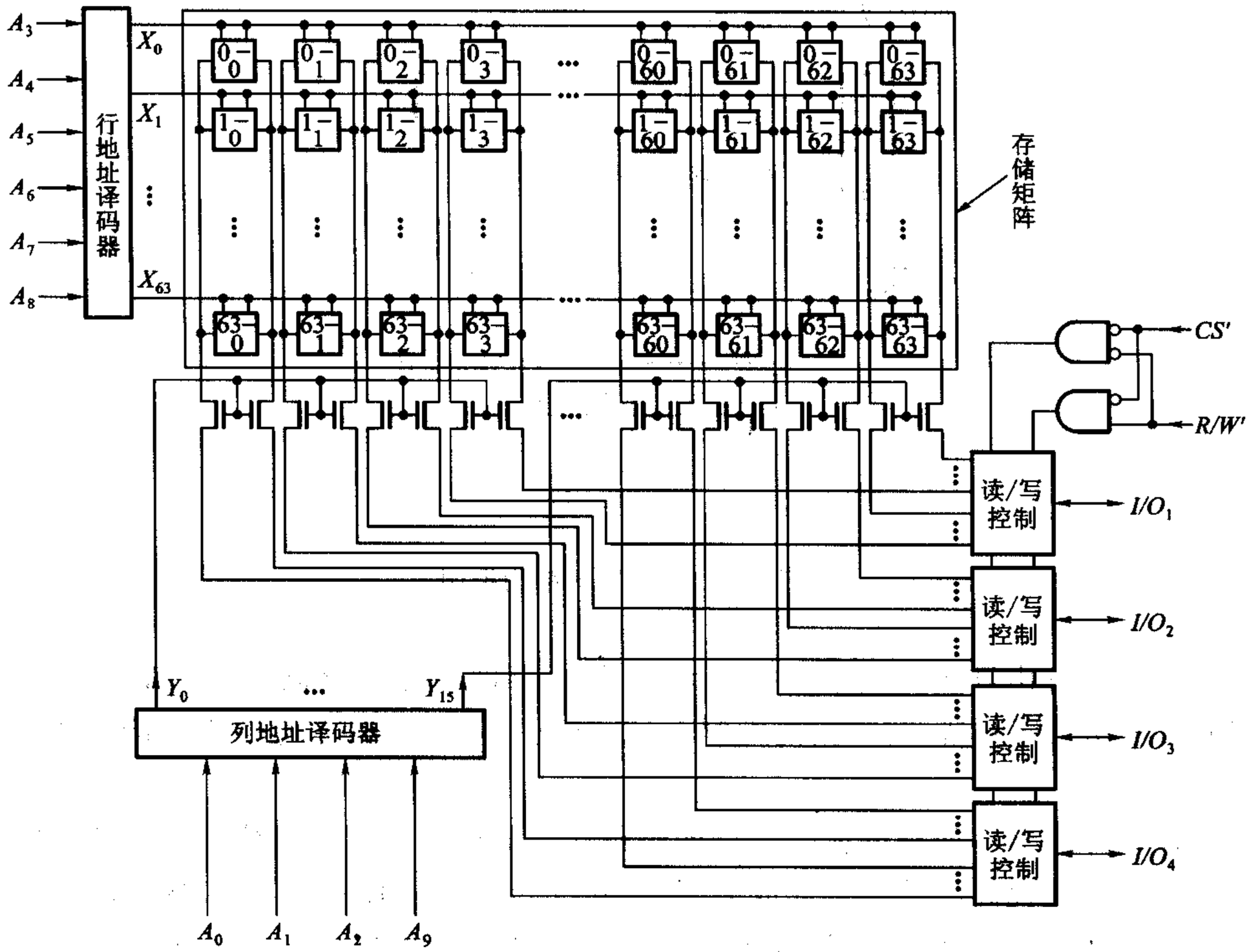


图 7.3.2　1024 × 4 位 RAM(2114)的结构框图

$I/O_1 \sim I/O_4$ 既是数据输入端又是数据输出端。读/写操作在 R/W'和 CS'信号的控制下进行。当 $CS'=0$,且 $R/W'=1$时,读/写控制电路工作在读出状态。这时由地址译码器选中的 4 个存储单元中的数据被送到 $I/O_1 \sim I/O_4$。

当 $CS'=0$,且 $R/W'=0$时,执行写入操作。这时读/写控制电路工作在写入工作状态,加到 $I/O_1 \sim I/O_4$ 端的输入数据便被写入指定的 4 个存储单元中去。

2114 采用高速 NMOS 工艺制作,使用单一的 +5 V 电源,全部输入、输出逻辑电平均与 TTL 电路兼容,完成一次读或写操作的时间为 100 ~ 200 ns。

若令 $CS'=1$,则所有的 I/O 端均处于禁止态,将存储器内部电路与外部连线隔离。因此,可以直接将 $I/O_1 \sim I/O_4$ 与系统总线相连,或将多片 2114 的输入/输出端并联运用。

二、SRAM 的静态存储单元

静态存储单元是在 SR 锁存器的基础上附加门控管而构成的。因此,它是靠锁存器的自保功能存储数据的。

图 7.3.3 是用六只 N 沟道增强型MOS管组成的静态存储单元。其中的 T_1 ~

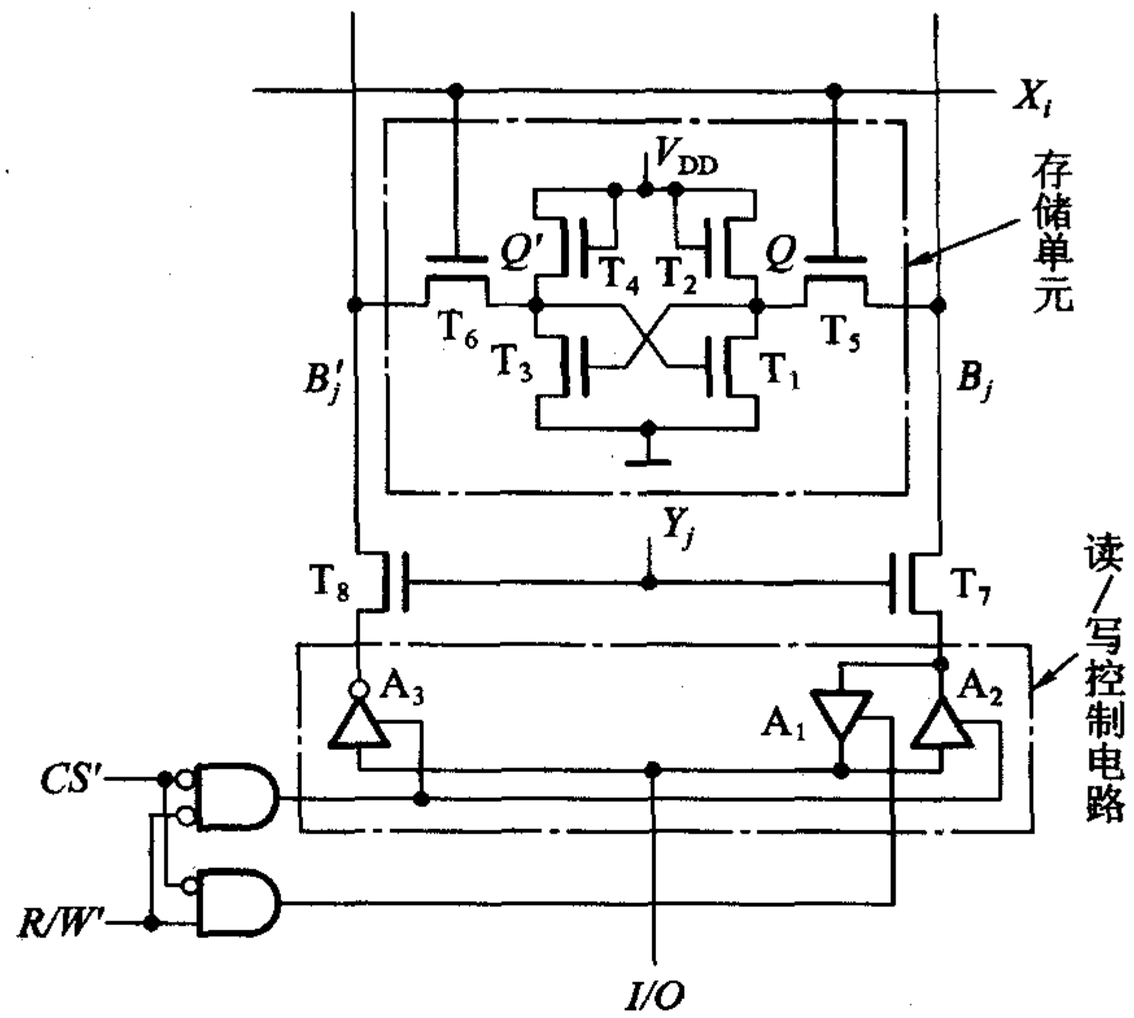


图 7.3.3 六管 NMOS 静态存储单元

T_4 组成 SR 锁存器,用于记忆 1 位二值代码。T_5 和 T_6 是门控管,作模拟开关使用,以控制锁存器的 Q、Q'和位线 B_j、B_j' 之间的联系。T_5、T_6 的开关状态由字线 X_i 的状态决定。$X_i=1$ 时 T_5、T_6 导通,锁存器的 Q 和 Q'端与位线 B_j、B_j' 接通;$X_i=0$ 时 T_5、T_6 截止,锁存器与位线之间的联系被切断。T_7、T_8 是每一列存储单元公用的两个门控管,用于和读/写缓冲放大器之间的连接。T_7、T_8 的开关状态由列地址译码器的输出 Y_j 来控制,$Y_j=1$ 时导通,$Y_j=0$ 时截止。

存储单元所在的一行和所在的一列同时被选中以后,$X_i=1$、$Y_j=1$,T_5、T_6、T_7、T_8 均处于导通状态。Q 和 Q'与 B_j 和 B_j' 接通。如果这时 $CS'=0$、$R/W'=1$,则读/写缓冲放大器的 A_1 接通、A_2 和 A_3 截止,Q 端的状态经 A_1 送到 I/O 端,实现数据读出。若此时 $CS'=0$、$R/W'=0$,则 A_1 截止、A_2 和 A_3 导通,加到 I/O 端的数据被写入存储单元中。

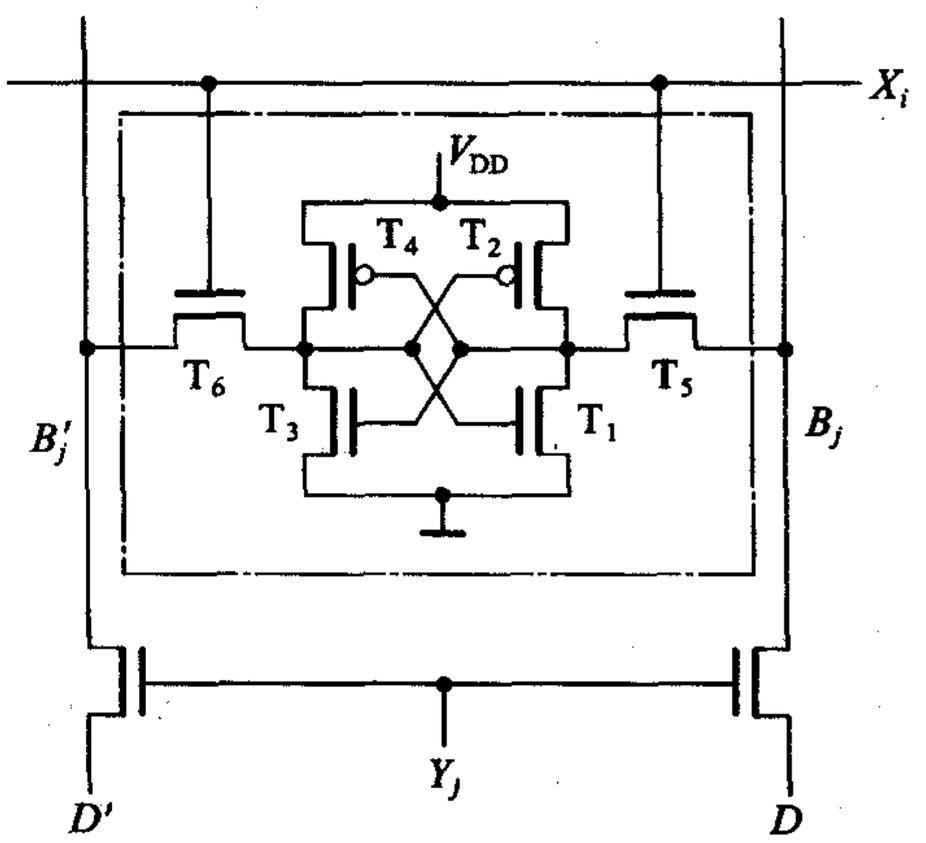


图 7.3.4 六管 CMOS 静态存储单元

由于 CMOS 电路具有微功耗的特点,尽管它的制造工艺比 NMOS 电路复杂,但在大容量的静态存储器中几乎都采用 CMOS 存储单元。图 7.3.4 是CMOS静态存储单元的电路,

它的结构形式与工作原理与图 7.3.3 相仿,所不同的是在 CMOS 静态存储单元中,两个反相器的负载管 T_2 和 T_4 改用了 P 沟道增强型MOS管。图中用栅极上的小圆圈表示 T_2、T_4 为 P 沟道MOS管,而栅极上没有小圆圈的为 N 沟道MOS管。

采用 CMOS 工艺的 SRAM 不仅正常工作时功耗很低,而且还能在降低电源电压的状态下保存数据,因此它可以在交流供电系统断电后用电池供电以继续保持存储器中的数据不致丢失,用这种方法弥补半导体随机存储器数据易失的缺点。例如,Intel 公司生产的超低功耗 CMOS 工艺的 SRAM5101L 用 +5 V 电源供电,静态功耗仅 1 ~2 μW。如果将电源电压降至 +2 V 使之处于低压保持状态,则功耗可降至 0.28 μW。

双极型 SRAM 的静态存储单元也有各种不同的电路结构形式,大体上分属于射极读/写存储单元、集电极读/写存储单元和集成注入逻辑存储单元三种类型。其中用得较多的是射极读/写存储单元,主要用在一些高速系统(如 ECL 系统)当中。这种存储单元的工作速度很快,但功耗较大。

图 7.3.5 是射极读/写存储单元的一个典型电路,它是由两个双极型多发射极三极管 T_1、T_2 和两个集电极负载电阻 R_1、R_2 组成的锁存器。一对发射极 e_{11}、e_{21} 与行地址译码器的输出线(字线)X 相连,另一对发射极 e_{12}、e_{22} 接到互补的位线 B 和 B' 上。电源电压 V_{CC} 通常取 3 ~3.5 V,位线的偏置电压 V_{BB} 约 1.4 V。

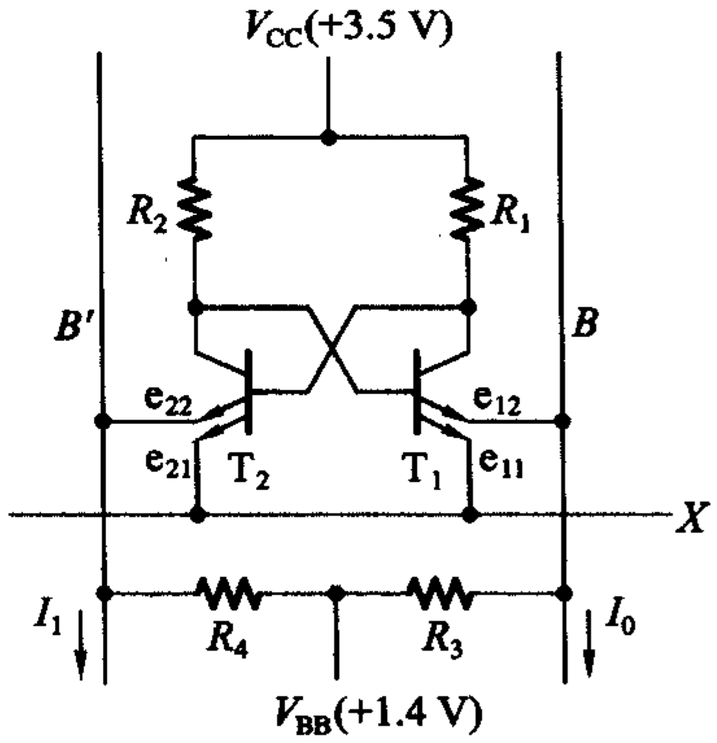


图 7.3.5　双极型 RAM 的静态存储单元

在保持状态下,字线 X 为低电平,低于 0.3 V,而位线 B 和 B' 为 1.4 V 左右,因此导通管的发射极电流由字线流出,而与位线相连的两个发射极 e_{12} 和 e_{22} 处于反向偏置状态,将存储单元与位线隔离。存储单元的状态可以是 T_1 截止、T_2 导通(定义为 **1** 状态),也可以是 T_1 导通、T_2 截止(定义为 **0** 状态)。

进行读出时,字线被提高至 +3 V,高于位线的 1.4 V,因而导通管的电流转而从位线流出。这时只需检测其中一根位线(B 或 B')上是否有电流流出,便能判断存储单元原来的状态是 **1** 还是 **0**。例如对位线 B' 的电流进行检测,如果存储单元为 **1** 状态,则 T_2 导通,B' 线在读出时有电流 I_1 流出。如果存储单元是 **0** 状态,则 T_1 导通,B' 线在读出时没有电流流出,$I_1 = 0$。将 B' 线上检测到的电流经读出放大器加以放大,就变成了高、低电平的输出电压信号了。

这种读出方式为非破坏性读出,即读出数据以后并不破坏存储单元原来的

状态。假定存储单元原为1状态,则 T_2 导通、T_1 截止。因此,当字线上升为3 V时 e_{11}和 e_{21}必然截止,而 e_{22}的电流流入位线 B'。若 T_2 导通后 c－e 间的压降小于0.3V,那么 c_2 的电位(也就是 b_1 的电位)将低于1.7V,不足以使 e_{12}导通,故 T_1 保持原来的截止状态。当字线回到低电平以后,流经 e_{22}的电流转向字线 X,并使 c_2 的电位跟随字线下降,锁存器继续保持1状态不变。

进行写入时首先也使字线变为 +3V,如果要求写入1,则加到存储器输入/输出端的信号经写入放大器转换后给出 $B=1$、$B'=0$ 的电压信号,使 T_1 截止、T_2 导通,锁存器被置1。反之,如果要求写入0,则给出 $B=0$、$B'=1$,将锁存器置0。

*7.3.2 动态随机存储器(DRAM)

一、DRAM 的动态存储单元

RAM 的动态存储单元是利用MOS管栅极电容可以存储电荷的原理制成的。由于存储单元的结构能做得非常简单,所以在大容量、高集成度的 RAM 中得到了普遍的应用。但由于栅极电容的容量很小(通常仅为几皮法),而漏电流又不可能绝对等于零,所以电荷保存的时间有限。为了及时补充漏掉的电荷以避免存储的信号丢失,必须定时地给栅极电容补充电荷,通常将这种操作称为刷新或再生。因此,DRAM 工作时必须辅以必要的刷新控制电路(控制电路通常是做在 DRAM 芯片内部的),同时也使操作复杂化了。尽管如此,DRAM 仍然是目前大容量 RAM 的主流产品。

早期采用的动态存储单元为四管电路或三管电路。这两种电路的优点是外围控制电路比较简单,读出信号也比较大,而缺点是电路结构仍不够简单,不利于提高集成度。单管动态存储单元是所有存储单元中电路结构最简单的一种。虽然它的外围控制电路比较复杂,但由于在提高集成度上所具有的优势,使它成为目前所有大容量 DRAM 首选的存储单元。

图7.3.6是单管动态MOS存储单元的电路结构图。存储单元由一只 N 沟道增强型MOS管 T 和一个电容 C_S 组成。

在进行写操作时,字线给出高电平,使 T 导通,位线上的数据便经过 T 被存入 C_S 中。

在进行读操作时,字线同样应给出高电平,并使 T 导通。这时 C_S 经 T 向位线上的电容 C_B 提供电荷,使位线获得读出的信号电平。设 C_S 上原来存有正电荷,电压 v_{C_S}为高电平,而位线电位 $v_B=0$,则执行读操作以后位线电平将上升为

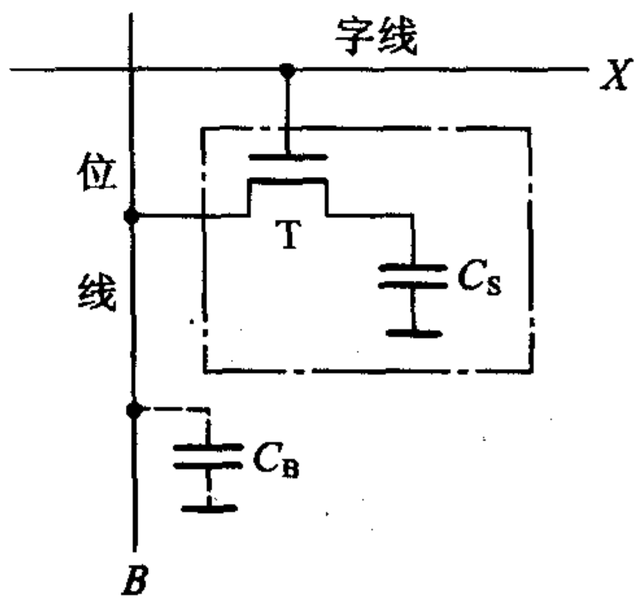


图7.3.6 单管动态MOS存储单元

$$v_B = \frac{C_S}{C_S + C_B} v_{C_S} \tag{7.3.1}$$

因为在实际的存储器电路中位线上总是同时接有很多存储单元，使 $C_B \gg C_S$，所以位线上读出的电压信号很小。

例如，读出操作以前 $v_{C_S} = 5$ V，$C_S/C_B = 1/50$，则位线上的读出信号将仅有 0.1 V。而且在读出以后 C_S 上的电压也只剩下 0.1 V，所以这是一种破坏性读出。因此，需要在 DRAM 中设置灵敏的读出放大器，一方面将读出信号加以放大，另一方面将存储单元里原来存储的信号恢复。

二、灵敏恢复/读出放大器

DRAM 中的单管动态存储单元也是按行、列排成矩阵式结构，并且在每根位线上接有灵敏恢复/读出放大器。图 7.3.7 是一个灵敏恢复/读出放大器的原理性电路图，它包含一个由 $T_1 \sim T_4$ 组成的锁存器和三个控制管 T_5、T_6 和 T_7。放大器的一个输出端与位线 B 和存储单元相连，另一个输出端接至一个虚单元上。虚单元的存储电容 C_F 上存入一个介于高、低电平之间的参考电平 V_R。

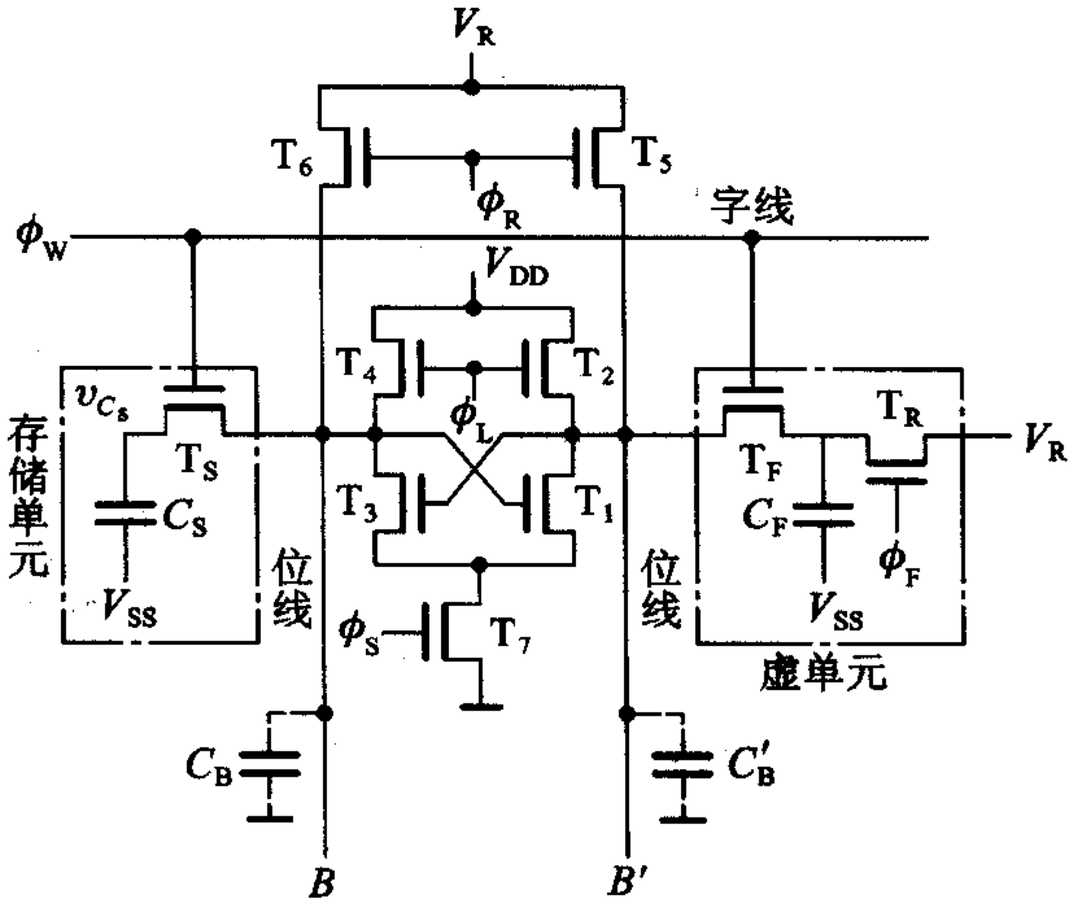
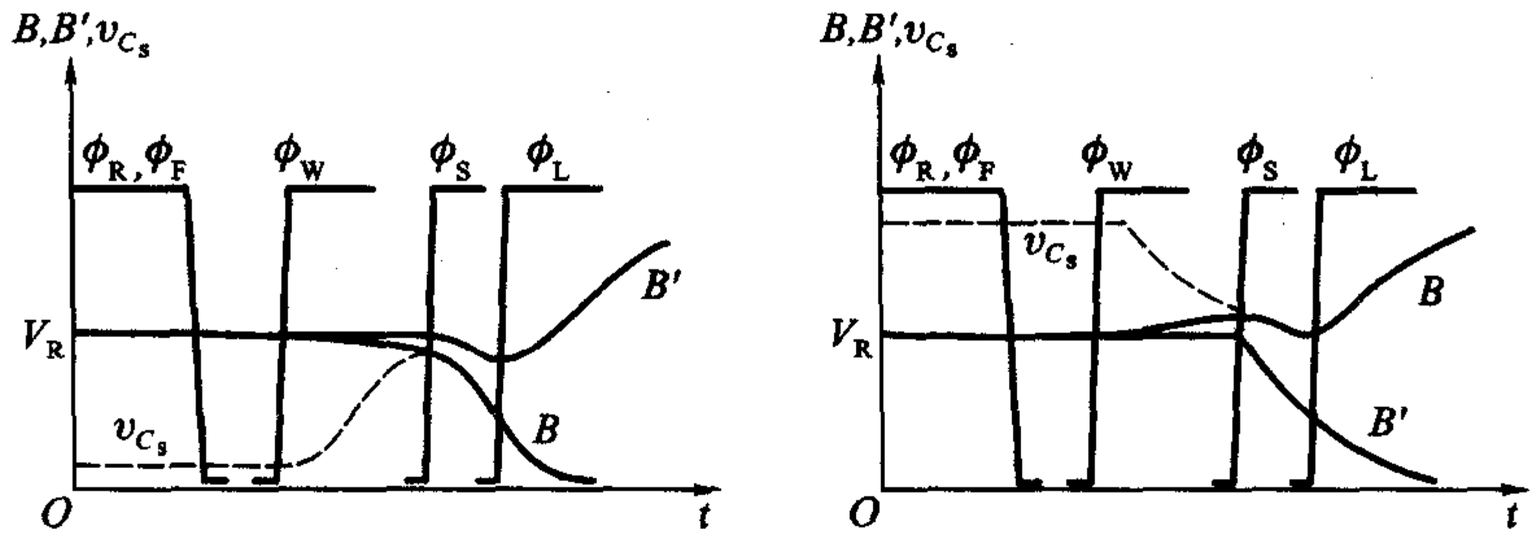


图 7.3.7　DRAM 中的灵敏恢复/读出放大器

图 7.3.8 中画了读出过程中 B、B'和 v_{C_S}的电压波形。读出过程在一组顺序产生的时钟信号控制下进行。首先 ϕ_R、ϕ_F 给出正脉冲，使 T_5、T_6、T_R 导通，位线 B、B'和 C_F 均被充电至 V_R。当字线选通脉冲 ϕ_W 到达后，存储单元的开关管 T_S 和虚单元的开关管 T_F 同时都导通。如果 C_S 上没有存储电荷（记忆的是 **0** 状态），则 C_B 经 T_S 向 C_S 放电，v_{C_S}上升而位线 B 的电位逐渐下降。当时钟信号 ϕ_S 到达后，位线 B 和 B'间的电位差被 T_1 和 T_3 组成的正反馈电路放大。最后 ϕ_L 脉冲使 T_2 和 T_4 导通，将 B'提升到正常的逻辑高电平，而 B 下降到正常的逻辑

低电平。在 ϕ_W 正脉冲消失以后，C_S 恢复为读出前的 **0** 状态。



(a) (b)

图 7.3.8 灵敏恢复/读出放大器的读出过程

(a) 读出 **0** 的情况 (b) 读出**1**的情况

如果 ϕ_W 正脉冲到达时 C_S 上存有电荷，v_{C_S}为高电平(记忆的是 **1** 状态)，则 C_S 经 T_S 向 C_B 放电，使位线 B 的电位上升。在 ϕ_S 正脉冲到达后，B 和 B'间的电位差被 T_1 和 T_3 组成的正反馈电路放大。然后在 ϕ_L 正脉冲到来后将 B 提升到正常的逻辑高电平而 B'降低到正常的逻辑低电平。ϕ_W 回到低电平以后 v_{C_S}保持高电平，C_S 恢复为读出前的 **1** 状态，所存储的 **1** 得到刷新。

由此可见，使用了图 7.3.7 所示的灵敏恢复/读出放大器之后，在每次读出数据的同时也完成了对存储单元原来所存数据的刷新。因此，DRAM 中的刷新操作是通过按行依次执行一次读操作来实现的。刷新时输出被置成高阻态。

三、DRAM 的总体结构

为了在提高集成度的同时减少器件引脚的数目，目前的大容量 DRAM 多半都采用 1 位输入、1 位输出和地址分时输入的方式。

图 7.3.9 是一个 64 K×1 位 DRAM 总体结构的框图。从总体上讲，它仍然包含存储矩阵、地址译码器和输入/输出电路三个组成部分。

存储矩阵中的单元仍按行、列排列。为了压缩地址译码器的规模，经常将存储矩阵划分为若干块。例如，图 7.3.9 的例子中是将存储矩阵划分为(1)、(2)两个 128 行、256 列的矩阵。

在采用地址分时输入的 DRAM 中，地址代码是分两次从同一组引脚输入的。分时操作由 RAS'和 CAS'两个时钟信号来控制。首先令 $RAS'=0$，输入地址代码的 $A_0 \sim A_7$ 位，然后令 $CAS'=0$，再输入地址代码的 $A_8 \sim A_{15}$位。$A_0 \sim A_6$ 被送到行地址译码器并被锁存，A_7 送入对应的寄存器。行地址译码器的输出同时从存储矩阵(1)和存储矩阵(2)中各选中一行存储单元，然后再由 A_7 通过输入/输