

目 录

第一章 电路形式和设计方法	1
1.1 MOS 晶体管的定性模型	1
1.2 静态逻辑形式	11
1.3 存贮元件及顺序电路	24
1.4 规则结构	32
1.5 微系统设计	39
1.6 小结	41
习题	42
参考文献	45
第二章 MOS 器件电子学	48
2.1 电容器和二极管	48
2.2 MOS 晶体管中的一级效应	58
2.3 MOS 晶体管中的二级效应	72
2.4 互连线的建模	91
2.5 散热	107
2.6 小结	108
习题	111
参考文献	113
第三章 芯片工艺	117
3.1 半导体器件的制造技术	117
3.2 两种基本的MOS加工工艺	125
3.3 设计规则	132
3.4 成品率	140
3.5 版图设计技巧	142
3.6 小结	144
习题	144
参考文献	145
第四章 逻辑抽象	148
4.1 逻辑电平的恢复	148
4.2 反相器	154
4.3 噪声	168
4.4 最坏条件设计	172
4.5 小结	181
习题	181
参考文献	183
第五章 电路技术	185
5.1 简单逻辑门的电路优化技术	185
5.2 大扇出电路和时钟驱动器	188

5.3 大扇入电路	199
5.4 双稳态	202
5.5 读出放大器	208
5.6 电平转换	215
5.7 输入保护	217
5.8 衬偏电压的产生	221
5.9 电路设计中的版图技术	227
5.10 小结	237
习题	237
参考文献	240
第六章 时钟与通讯	243
6.1 单相时钟定时	243
6.2 两相时钟定时	244
6.3 重叠时钟	258
6.4 自定时 MOS 电路	261
6.5 同步器和判断器	264
6.6 全局时钟定时问题	268
6.7 小结	272
习题	273
参考文献	275
第七章 阵列结构的电路技术	276
7.1 译码器	276
7.2 ROM 和 PLA	279
7.3 RAM	284
7.4 收缩阵列	292
7.5 小结	299
习题	300
参考文献	302
第八章 超大规模集成系统的微结构	305
8.1 数据通道设计	306
8.2 控制结构的实现	313
8.3 测试和故障排除	322
8.4 小结	325
习题	326
参考文献	326
附录	330
A. SPICE 技术	330
B. 非线性电容方程	330
C. 2 μ m nMOS 工艺范例	331
D. 2 μ m CMOS 工艺范例	335
E. nMOS 和 CMOS 工艺的两项先进备选工艺	338
F. 有用的常数和换算	338
索引	340

第一章 电路形式和设计方法

集成电路设计所涉及的专业范畴从器件物理、非线性电路理论一直延伸到数字系统微结构。在 VLSI 技术中, 集成电路设计是核心, 它包括了不同层次的抽象: 静电学、拓扑学、开关理论等等。VLSI 设计者的任务是如何超越这些截然不同的领域并把一种设想转变为一个制作电子设备的技术细节。

每一抽象在细节上都有一个层次结构。系统必须分解为子系统, 子系统再分解为模块, 模块分解为元件。随着每一步分解, 电子设备也就变得更加具体化。对抽象和细节进行二维分层的方法有多种, 但不论选取什么方法, 它的最后成功都取决于自下而上地对各个层次的理解。由于这一原因, 我们从晶体管和互连线这个较低的层次上开始对 VLSI 设计的研究。

1.1 MOS 晶体管的定性模型

在本章中, 我们的目的是定性地了解 MOS 数字电路是如何工作的。作为这一任务的一部分, 我们需要了解一些 VLSI 中的低层次元件的行为以及它们如何组成有意思的功能块。我们将讨论几种不同的设计方法, 但应该把它们看作是更大范围内的统一整

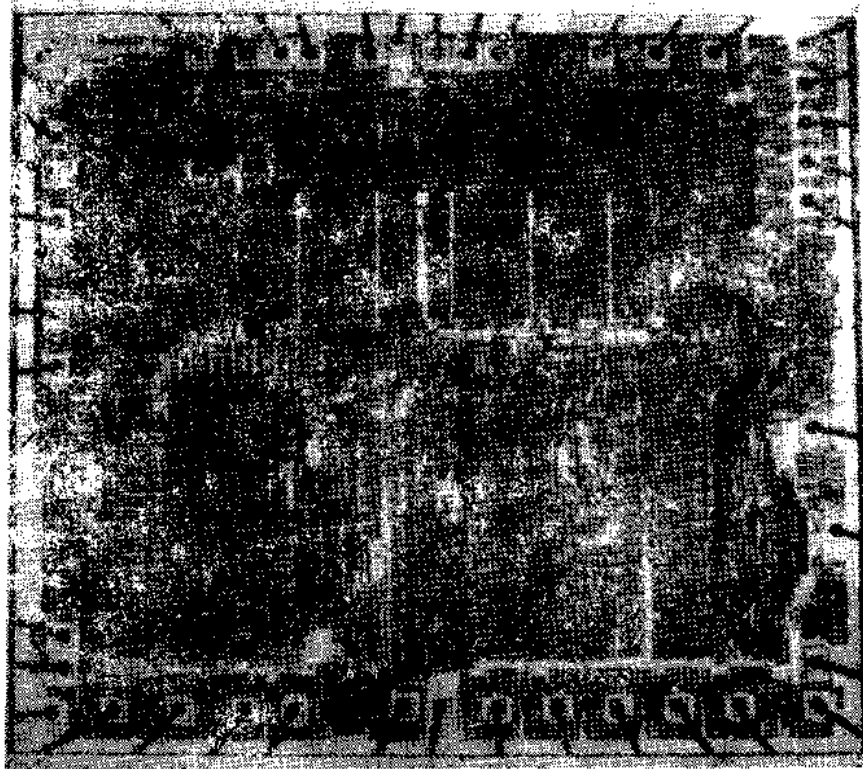


图1.1 T11微处理器的显微照片

体的一部分。在后面的章节中，还将对这些以及其它一些技术作更精密的定量讨论。

图 1.1 为一个 LSI 芯片的显微照片 [Olsen 81]。集成电路的二维性是明显的。在今天的各种 VLSI 中只用了两种基本的元件——晶体管与导线。在图 1.1 中我们见到的大部分是连线，晶体管所占的面积仅为 10%。这一芯片包含了 13000 个晶体管（按现代的标准，这只是个小芯片），它的边长为几毫米，厚度为几分之一毫米，它的主要结构都在表面以下几微米深度的范围内。有几个大的规则结构很显眼，大多数晶体管都在这几个规则结构中。但就工作量而言，它们仅占几分之一。为了看到单个的晶体管，需要更细致地观察芯片。

晶体管位于两条导线交叉的位置。图 1.2 为部分电路的特写顶视图，不过，我们看到的大部分仍是连线，只有知道晶体管如何找的人才能看到晶体管。由此可以明白，为什么常常说在 VLSI 中晶体管是“白送的”，主要的花费是连线。在这一具体芯片中有三层互连线材料。有斑点部分的材料是铝。另外两种连接材料是多晶硅和重掺杂单晶硅。这两种材料与晶体管的关系很和谐，下面我们将会看到这一点。

图 1.3 为图 1.2 所示电路的版图。各种材料都在图中标出。多晶硅（多晶）和金属层都是在制造芯片的工艺过程中淀积上去的。金属只起连线的作用而多晶则起着连线和

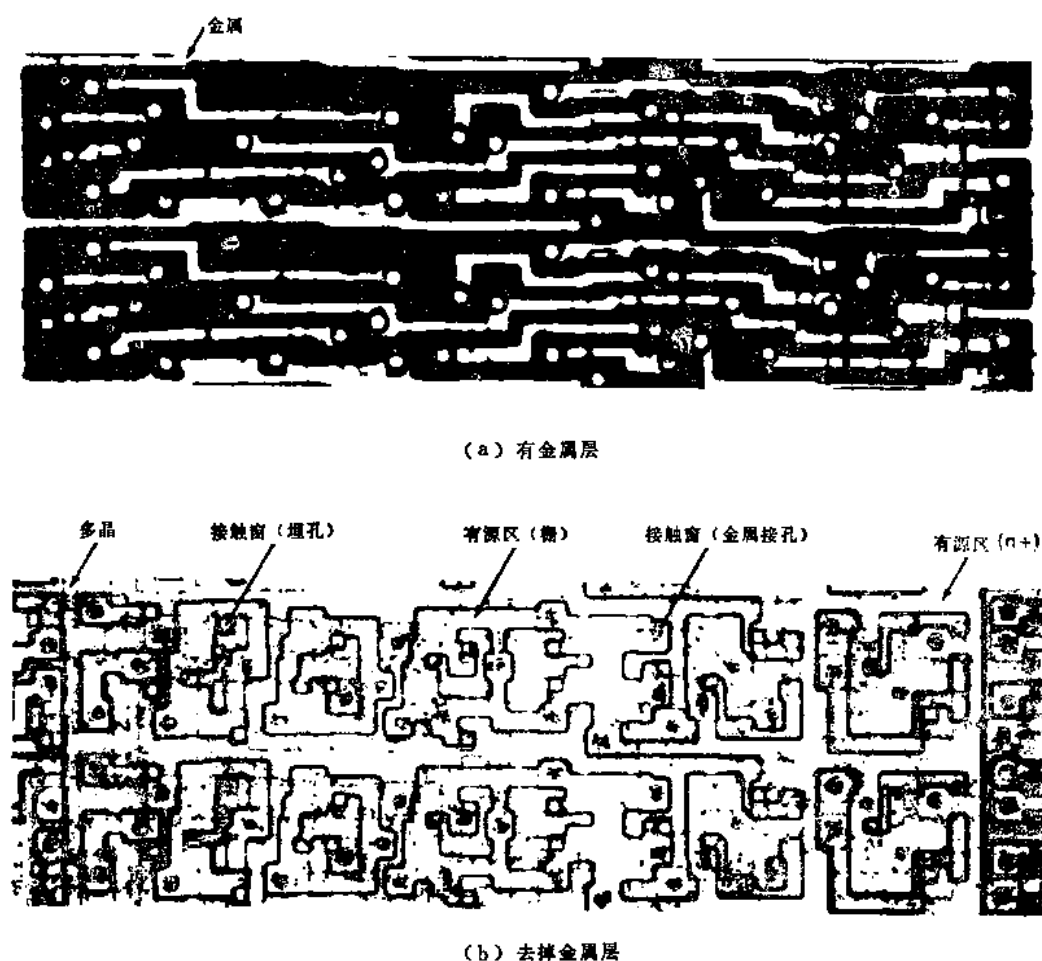


图1.2 T11微处理器的局部放大显微照片

晶体管栅电极的双重作用。有源区，包括重掺杂的单晶硅区和晶体管的栅区，也起着双重作用。有源区被多晶分为两部分，在没有多晶的有源区，硅掺杂为 n^+ 型， n^+ 材料起导线作用。在多晶与有源区的交叉处形成了晶体管。二氧化硅（玻璃）层把三个导电层分开。这些氧化层之间略有不同（对此将在第三章讨论），它们包括栅氧化层，场氧化层，CVD 氧化层和表面覆盖玻璃层。为了连接两种不同的导电层，用了两种类型的接触孔。第一种接触孔把金属与多晶或 n^+ 连在一起，第二种接触孔称为埋孔，它把多晶与 n^+ 连在一起。图 1.4 所示为一种典型互连结构的剖面图。若在显微镜下观察，可以看到一彩色条块拼凑图。多晶可能是红的而 n^+ 是绿的或相反。这些颜色大部分都是“假的”，它们是由光波在不同结构上的干涉而形成的，这些结构的厚度与光波长同量级。

MOS¹⁾ 晶体管是一种四端器件。在正常的工作条件下，晶体管的栅控制源漏²⁾ 之间的载流子电流。第四个端子是衬底，在 pMOS 及 nMOS 技术中，不能用衬底完成逻辑功能，这是因为在 nMOS 技术中，所有衬底端都是连在一连的。图 1.5 所示为一个 nMOS 晶体管的截面。源及漏由重掺杂的 n^+ 硅组成，有一层厚度仅为几百个原子的二氧化硅层把栅与沟道区隔开，栅导体是用重掺杂的多晶硅做成的。对这些几何形状给出一个清晰的图象是必要的。图 1.6 为一个 nMOS 晶体管的三视投影图。图 1.7 所示为在一个共同衬底上的两个 nMOS 器件。这些器件的源和漏是由反向偏置的 p-n 结及与之相联系的耗尽区互相隔离开的（耗尽区是半导体中缺乏载流子的区域）。

有两种类型的 nMOS 晶体管：常通型与常断型。在常断型 nMOS 晶体管中，栅电压为 0 时，栅下的沟道区中没有可动载流子。这种类型的晶体管称为增强型器件，这是因为这种晶体管在零栅压下，硅表面处于增强状态。

- 1) MOS 为 Metal Oxide Semiconductor 的缩写，表示器件的物理结构，MOS 也称为 IGFET (Insulated Gate Field Effect Transistor)，CMOS，pMOS，nMOS 分别代表互补、P 沟、n 沟 MOS。对于现代 MOS 器件，M 并不一定是金属栅，而往往用硅栅取而代之。
- 2) 之所以称之为“源”，是因为它是可动载流子的源泉。对 n 沟道器件，这些载流子为电子；对 P 沟道器件，它们是空穴。

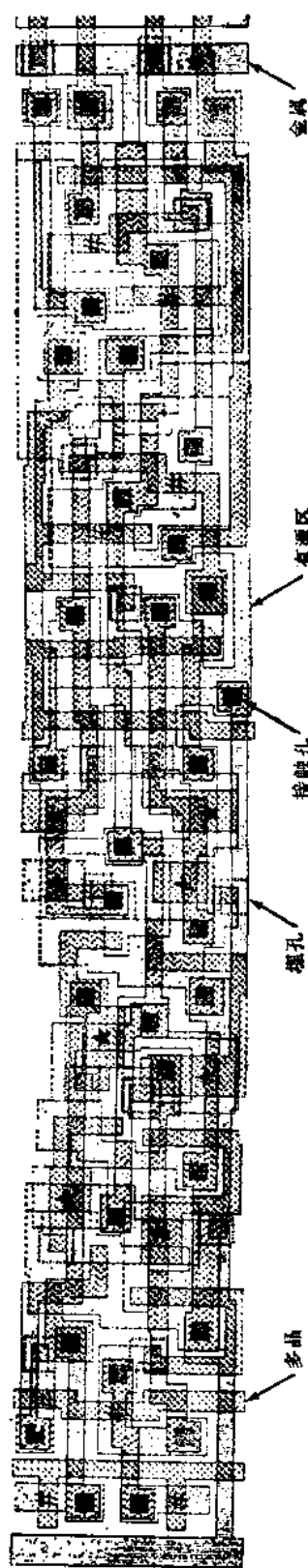


图 1.3 图 1.2 所示部分的版图

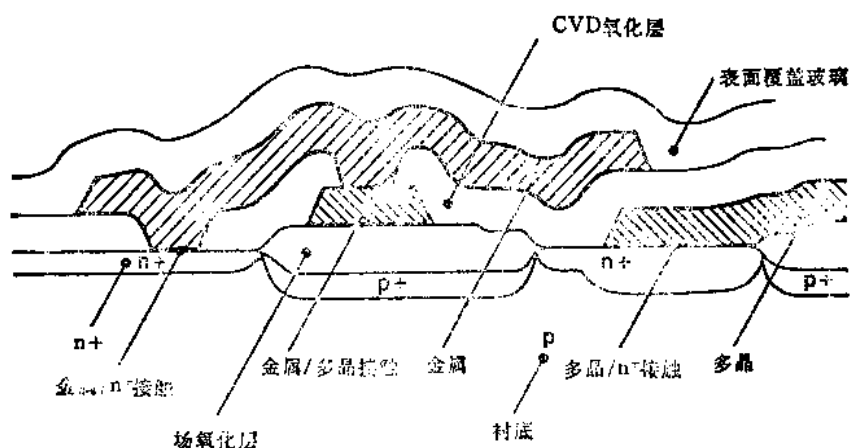


图1.4 nMOS 工艺中的一些接触孔

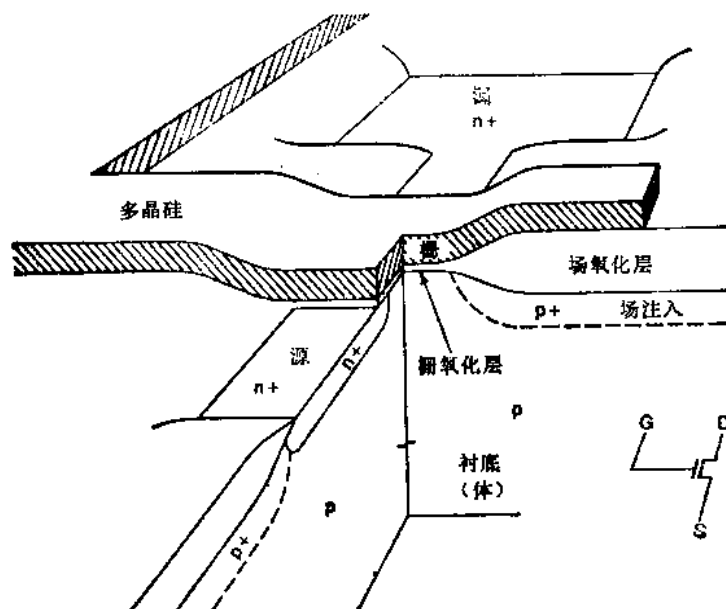


图1.5 工艺过程中经源漏注入后的n沟MOS晶体管的新画圈

当在nMOS晶体管的栅上加一正电荷 Q_g 时，在沟道区将感应出几乎等于 $-Q_g$ 的电荷。这一电荷由耗尽的受主离子及流动于源漏之间的自由载流子组成。假定把晶体管设计成在栅被充电到电源电压 V_{DD} 时沟道电荷主要是可动载流子，即电子。在漏对源加正电压 v_{DS} 将使电子由源流出而被扫入漏，电子从源流到漏需要的时间为渡越时间 t_t 。它取决于沟道载流子的迁移率 μ 、电场强度 $\mathcal{E}$ （它使载流子移动），和沟道长度 L 。每个渡越时间之后，沟道电荷将完全更新。迁移率说明载流子平均漂移速度 v 与电场的关系：

$$v = \mu \mathcal{E} \quad (1.1)$$

电流 i_{DS} 等于 Q_g/t_t 。在某一电流值之下，源漏电压 v_{DS} 越高，源漏电流越大，理由是，较高的源漏电压将使沟道中的电场较大，从而使载流子的运动加快。当栅电压 v_{GS} 为低电

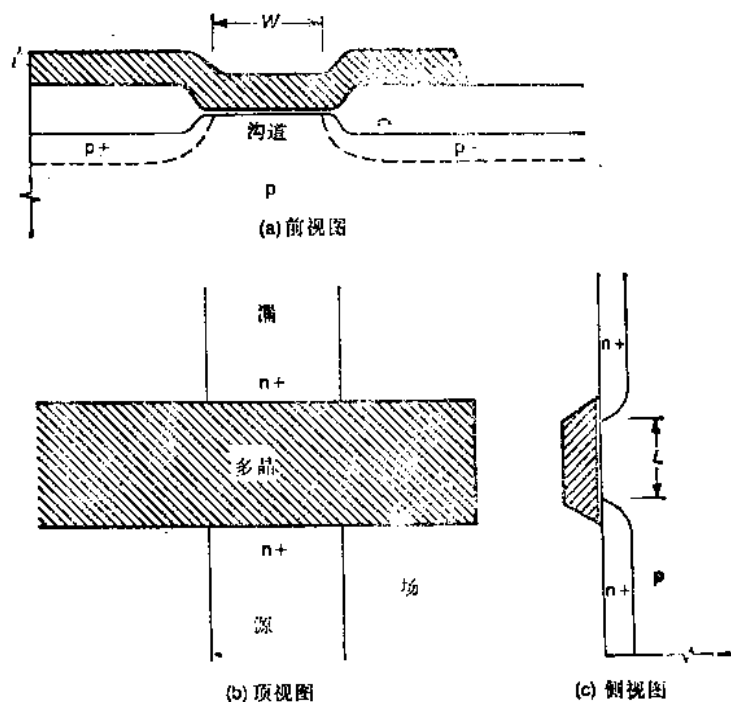


图1.6 n沟MOS晶体管的三视图

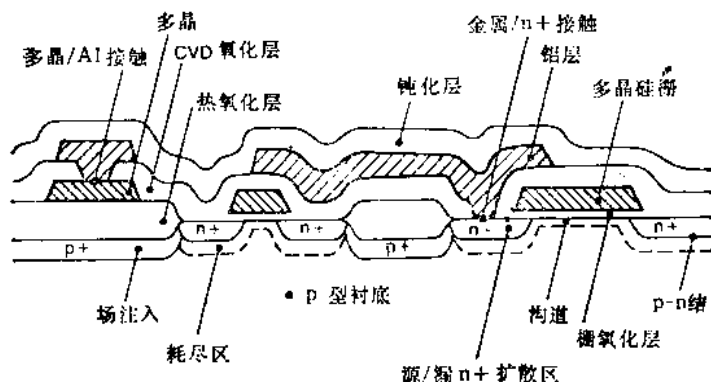


图1.7 p型衬底上的两个n MOS晶体管。晶体管之间的隔离是靠反向偏置的p-n结

平(Q_c 小)时,在沟道中很少或没有可动电荷。刚能在沟道中感应出可动电荷的栅电压称为阈电压 V_T 。当栅-源电压低于 V_T 时,源漏之间的电阻接近于无穷。对于高于 V_T 的栅电压,典型的MOS晶体管的电阻在 $20k\Omega$ 的量级。栅电压越高,有效电阻 R 越低,因为这时在沟道中将感应出更多的可动电荷,使导电的效率增加。所以对一定的漏源电压来说,栅电压越高漏电流就越高。粗略地说,栅电压高于阈值时,

$$R \propto \frac{1}{v_{GS} - V_T} \quad (1.2)$$

典型的增强型nMOS晶体管的电流-电压特性如图1.8所示。

另一种说明MOS晶体管原理的方法是借助于电位。图1.9所示为在正的漏-源电

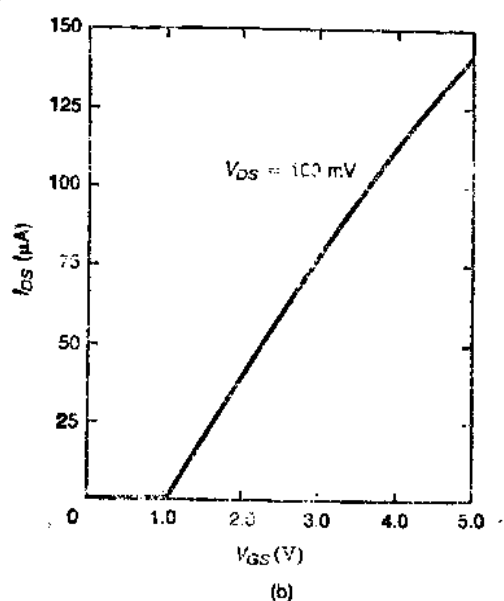
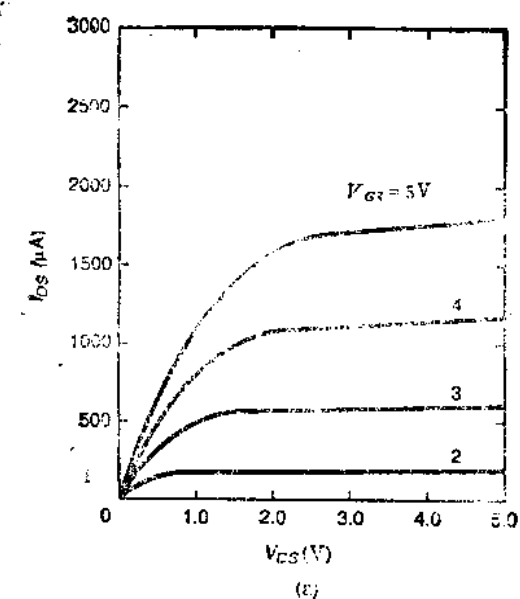


图 1.8 n 沟道MOS器件的电流-电压特性
(a) 漏电流与源漏电压的关系;
(b) 漏电流与栅源电压的关系

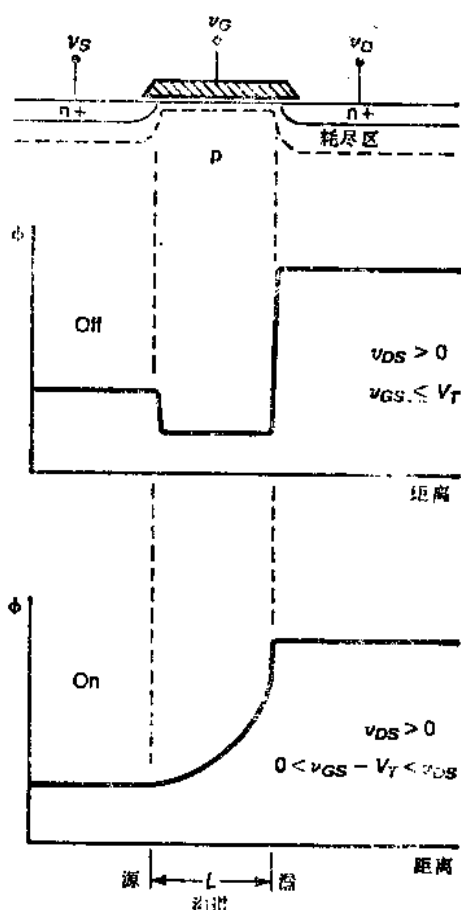


图 1.9 MOS晶体管栅下的电位分布
电子被吸引上电位坡达到正的漏电压

压¹⁾下增强型nMOS晶体管沿沟道的电位分布。我们看到当器件工作在阈电压以下时源处的势垒将阻止电子流动。这一势垒是由n型源与p型沟道区之间的p-n结自建势形成的。正的栅电压将拉高栅下的电势从而降低势垒，达到阈值时，势垒将变得很低使电子可以流动。栅电压越高，沟道中的电子就越多。一旦进入沟道，电子就被侧向电场所加速，沿电势坡上爬（沿势垒下落）。当栅电压大于阈电压时，尽管材料是p型的，沟

1) 我们把图安排为，被正的漏电位吸引的电子需要爬坡流动。

道中的载流子仍为电子。换句话说,当晶体管导通时,沟道为有效n型。这就是为什么沟道区为p型材料的晶体管称为n沟道晶体管的原因。

到此为止,我们所讨论的仅仅是正常情况下不导通的晶体管。由于我们仅用晶体管及连线组成集成电路,当我们需要一个电阻时,就必须用晶体管来构成它。一个常通晶体管可以像某种非线性电阻一样工作,理论上,制作一个称之为耗尽型晶体管的常通晶体管是容易的。为了使增强型晶体管导通,在栅上加电荷,使沟道区感生出镜像电荷。如果要使晶体管永远导通,可以在沟道区加固定电荷,这些电荷的作用就像一个与栅串联的电池。我们仍然可以把这一器件关断,不过这时加的栅压必须把沟道载流子耗尽。要注意,即使是 $v_{GS}=0$,栅下的区域也仍然为有效n型。

除了耗尽型和增强型n沟晶体管之外,也可以采用p沟晶体管。在nMOS中我们只用n沟晶体管,在pMOS中我们只用p沟晶体管,但在CMOS中二者都有。p沟晶体管的工作情况刚好与n沟器件相反。对于增强型pMOS器件,当 v_{GS} 为0时晶体管是不导通的(正如nMOS的情形一样),但栅压为负时导通。图1.10所示为一个p沟增强型晶体管的电流-电压特性。

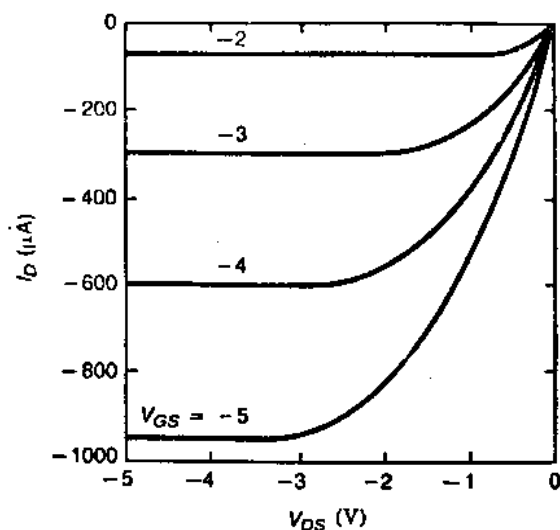


图1.10 p沟晶体管的电流-电压特性

MOS晶体管的“体”(衬底)端对其电学特性有重要的(但却是缓和的)影响。在正常工作中,nMOS器件的衬底相对于源和漏端不能是正确的,以保证隔离的p-n结反向偏置。对于pMOS晶体管,衬底对源和漏区不能是负的¹⁾。

当衬底电压改变时,晶体管的电流-电压特性也有轻度改变。我们考察一个nMOS晶体管。假定晶体管的源端接地,而 v_s 为负值,当 v_s 变负时,栅下的势垒升高。为了使晶体管导通,栅电压必需更高,以克服增高的势垒。从模型的角度,这意味着 v_s 变负将使器件的阈电压升高。观察这一问题的另一种方式是,一部分栅电荷原先由沟道可动电子抵消,而现在则由衬底中加大的耗尽区电荷抵消。为了在沟道区获得同样数目的载流子就需要更多的栅电荷。我们可以把栅与衬底端都看作是控制沟道电荷的杠杆。在一个设计良好的器件中,栅端有大得多的控制作用。

p沟和n沟晶体管的IEEE标准符号[IEEE78]如图1.11(a)所示。图1.11(b)所示为在本书中将采用的四种不同类型晶体管的符号,它们包括三种nMOS器件(增强型,耗尽型和零阈值器件)和一种增强型pMOS器件。注意在这种记号中不画出衬底端。零阈值器件正如其名字所表示的那样,是一种阈电压做得接近于0的MOS晶体管。这些

1) 原文此后有一小段有误,译时略去。——译者注

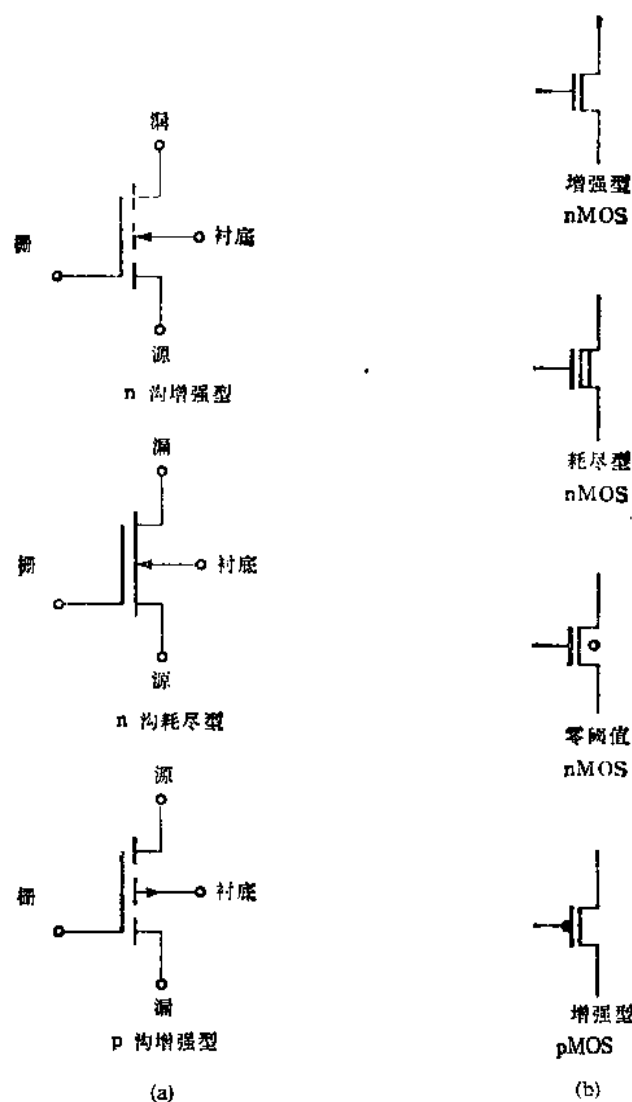


图1.11 MOS 晶体管的符号。(a) IEEE符号；(b) 本书使用的符号

器件的阈值一般为十分之几伏。它们也曾被称为“自然”晶体管，这是因为在某些工艺中这种器件可以被简单地“自然”做出来，即需要的工艺步骤最少。pMOS 器件符号上的圈可以被设想为栅上信号的逻辑反转。对于 pMOS 器件，栅上的高电平使晶体管关断而不是像 nMOS 情形那样使晶体管导通。对于工作电压为 +5 V 的工艺设计，增强型 nMOS 器件的阈值一般取 0.6 到 1.1 V，而 pMOS 器件一般取 -0.6 到 -1.1 V。耗尽型 nMOS 的阈值一般在 -2.5 到 -4 V 之间，而零阈值器件的阈电压一般在 0.2 到 0.6 V 之间。

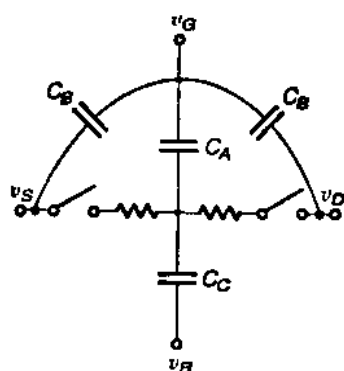


图1.12 MOS晶体管的电容。
理想晶体管只有电容 C_A

由于一个很宽的晶体管与并联起来的几个较窄的晶体管是一样的，因而MOS晶体管的导电能力随晶体管的有效宽度 W 的增加而增加，随晶体管有效沟道长度 L 的

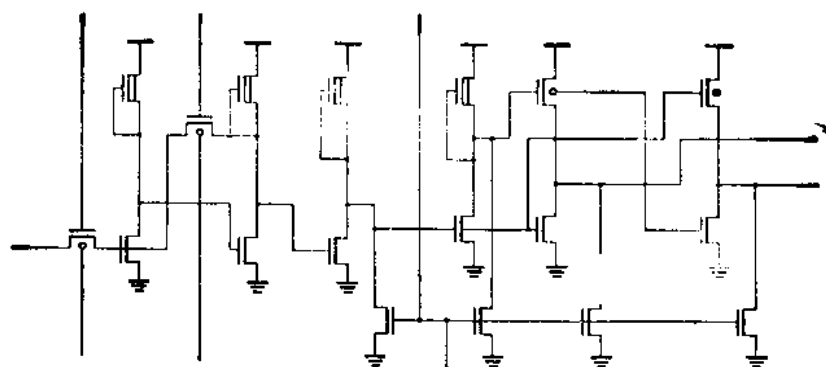


图1.13 图 1.2中电路的电路图

增加而减小。把 MOS 晶体管设想为一个电压控制的非线性电阻是有帮助的，电阻与长度成正比而与宽度成反比。我们定义一个形状因子 S ：

$$S \equiv \frac{W}{L} \quad (1.3)$$

粗略地说， S 与晶体管的电导成正比。

晶体管除了有一定的阈电压和电阻之外，还有多种电容。图 1.12 所示为 MOS 晶体管的简单模型，在阈值以下，开关是打开的。在阈值以上，源漏之间的非线性电阻起作用。 C_A 为栅与沟道之间的二氧化硅平板电容器电容。 C_B 为栅与源及栅与漏间的覆盖电容。 C_B 是重要的，即使晶体管开路，它也存在。 C_C 为沟道与衬底之间的电容。在典型的 MOS 器件中， C_A ， C_B ， C_C 的比例约为 40:5:1。

图 1.13 是按图 1.2 所示芯片图形画出的电路图。在 VLSI 设计中，我们试着画出表明芯片版图的电路图，这是由于连线对决定 VLSI 芯片性能有重要作用。为了实际做出一个芯片，我们需要产生一定的芯片图形，这些图形称之为掩模图形或版图。图 1.13 所示的布局就可以用来制作图 1.2 所示电路。

在体 CMOS 设计中，我们把 n 沟和 p 沟器件中的一种置于阱中，从而把它们分割开来。例如，在 n 阱工艺中， p 沟晶体管是位于阱中的。图 1.14 画出了一个在阱中的晶体管的剖面图。图 1.15 所示为一个 CMOS 芯片的部分显微照片 [Rubinfeld 82]。

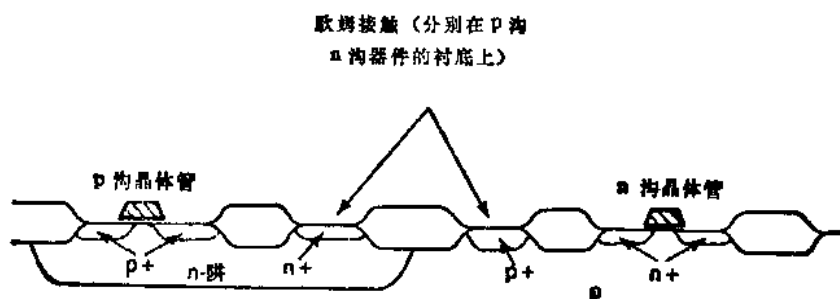


图1.14 n 阱CMOS晶体管的剖面

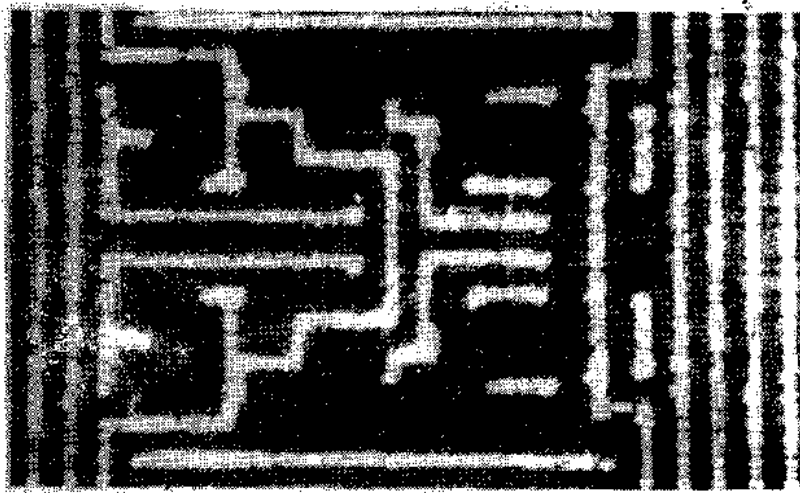


图1.15 J11微处理器的显微照片。采取双层多晶P阱CMOS工艺

例1.1 工艺范例。

本书从头至尾采用了两种人为的工艺，一种 nMOS 工艺和一种 CMOS 工艺，作为范例并作为习题的基础。尽管这些工艺并不代表任何一家公司的工艺，但它们的参数是合理的。该工艺是 $2\mu\text{m}$ 工艺，这意味着多晶硅栅的最小长度为 $2\mu\text{m}$ ¹⁾。设计规则、工艺参数、最坏条件 SPICE 文件等均列于书后附录中。SPICE [Vladimireson 80, 81] 是被

表1.1 用以产生图1.8 的 SPICE 输入文件

```

I-V characteristics for SS model of CMOS devices
*****
.MODEL NSS NMOS LEVEL=3 RSH=0 TOX=275E-10 LD=0.1E-6 XJ=0.14E-6
+ CJ=1.6E-4 CJSW=1.8E-10 UO=550 VTO=1.022 CGSO=1.3E-10
+ CGDO=1.3E-10 NSUB=4E15 NFS=1E10
+ VMAX=12E4 PB=0.7 MJ=0.5 MJSW=0.3 THETA=0.06 KAPPA=0.4 ETA=0.14
.MODEL PSS PMOS LEVEL=3 RSH=0 TOX=275E-10 LD=0.3E-6 XJ=0.42E-6
+ CJ=7.7E-4 CJSW=5.4E-10 UO=180 VTO=-1.048 CGSO=4E-10
+ CGDO=4E-10 TPG=-1 NSUB=7E15 NFS=1E10
+ VMAX=12E4 PB=0.7 MJ=0.5 MJSW=0.3 ETA=0.06 THETA=0.03 KAPPA=0.4
*****
M1 1 10 0 0 NSS W=13.2U L=2.25U
VDS 20 0
*****
* VGS is positive for nMOS and negative for pMOS
*****
VGS 10 0 5V
VIDS defines current direction for drain
*****
VIDS 20 1
.DC VDS 0 5 0.05
.PRINT DC I (VIDS)
.PLOT DC I (VIDS)
.WIDTH IN=75 OUT=75
.END

```

1) 关于 $2\mu\text{m}$ 工艺中究竟哪一种重要尺寸为 $2\mu\text{m}$ 并没有普遍承认的规定。此外也请注意， $2\mu\text{m}$ 工艺的 Mead and Convey 参数 λ 为 $1\mu\text{m}$ [Mead 80]。

广泛应用的、由加州大学伯克利分校开发的一种电路模拟程序；图 1.8 和图 1.10 所示的晶体管曲线是由 SPICE 程序 CMOS 器件模型产生的。所画器件尺寸为 $15/2$ ，这表示电路设计者所定的晶体管宽度为 $15\mu\text{m}$ ，长度为 $2\mu\text{m}$ 。表 1.1 所列用于产生其中一组曲线的 SPICE 输入文件。

1.2 静态逻辑形式

在集成电路芯片上，用晶体管和互连线可以做许多种不同部件，如运算放大器、收音机、存储器，数字开关网络，以及众多的高档电子产品等。我们的主要兴趣在于数字开关网络，我们将研究的最重要的数字电路元件是可恢复的逻辑门。逻辑门通常是不可逆的，它们将信号由输入端传到输出端但不能反过来。它们也是用以衰减噪声干扰效应的有源器件，目的是保存和刷新（恢复）完整的逻辑信号。

另一种有用的开关网络形式是所谓的传输晶体管，或继电器，或导引逻辑〔Shannon 37〕。在传输晶体管逻辑中，传输器件作为一种开关，有条件地把两个节点连在一起。当逻辑功能可方便地概括为把信号或标志在一定条件下经过一个网络加以导引时，往往适于用传输晶体管逻辑。图 1.16 所示为选择电路的两种形式，一种是用传统的复位逻辑门构成的，另一种是用传输晶体管。对传输晶体管的方案，从未标明的电路图中看不出哪个是输入端，哪个是输出端。实际上，信号可以向不同方向传输，这取决于外电路。在传输晶体管网络中，信号传输方向与设计者的意图相反是很常见的，所以这样的传输通道有“暗道”之称。

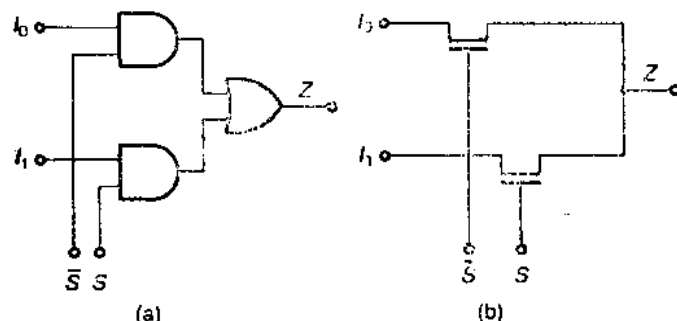


图1.16 两种 2 对 1 选择器电路

(a) 由门电路组成；(b) 由传输 nMOS 晶体管组成

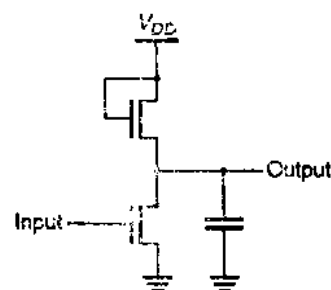


图1.17 nMOS全增强型反相器

全增强型 nMOS

所有 MOS 逻辑门的基础都是反相器。nMOS 反相器需要两个晶体管，一个晶体管在输入为低电平时把输出拉向高电平，一个晶体管在输入为高电平时把输出拉向地。在静态 nMOS 电路中，“上拉”晶体管通常接成二端器件。它的电阻与输入信号仅有间接关系。这一晶体管被称为“负载”。“下拉”晶体管的栅通常接到输入端。静态 nMOS 逻辑门是一种有比逻辑，所以如此称呼是因为它的一个逻辑电平取决于两个电阻的比。

最简单的反相器包括两个增强型器件。图 1.17 所示的增强型负载反相器工作原理

如下。当输入电压在 V_T 以下时，下拉晶体管不通，输出电压升至 $V_{DD}-V_T$ 。注意输出电压并不一直升到电源电压 V_{DD} ，这是因为一旦上拉晶体管的源升到 $V_{DD}-V_T$ 时，栅对源的电压降到 V_T ，于是晶体管关断。这时两个晶体管都不通，由于输出端电容的缘故，输出停留在 $V_{DD}-V_T$ ¹⁾。

当输入为高电平时，两个晶体管都导通，输出电压由两导通器件的分压决定。若上拉和下拉晶体管的形状因子之比设计得正确，当输入电压升到接近于 $V_{DD}-V_T$ 时，输出电压将降到 V_T 以下，这就导致一种自洽的情况，输入为高电平时输出为低电平，反之亦然。这就是一个反相器。

以反相器为基础，我们可以开始设计更复杂的逻辑门。图 1.18 所示为一个 3 输入端或非门，这个门含有一个上拉晶体管 and 三个并联的下拉晶体管，任何一个下拉器件导通，输出就被下拉到低电平。可以指出，实现或非门的能力可以用于实现任何逻辑功能。当然，这并不是说它是特别方便的或有效的实现方法。图 1.19 所示为用或非门实现的一个全加器。全加器满足下述方程：

$$S = A\bar{B}\bar{C} + \bar{A}B\bar{C} + \bar{A}\bar{B}C + ABC \quad (1.4)$$

$$C_o = AB + BC + AC \quad (1.5)$$

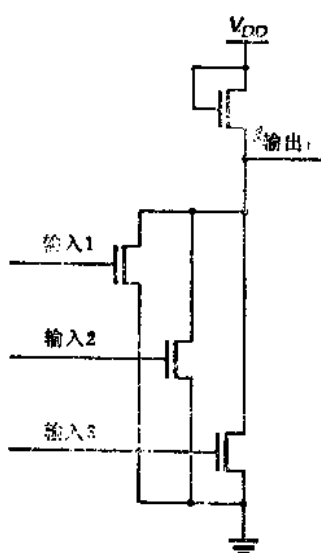


图 1.18 3 输入端或非门

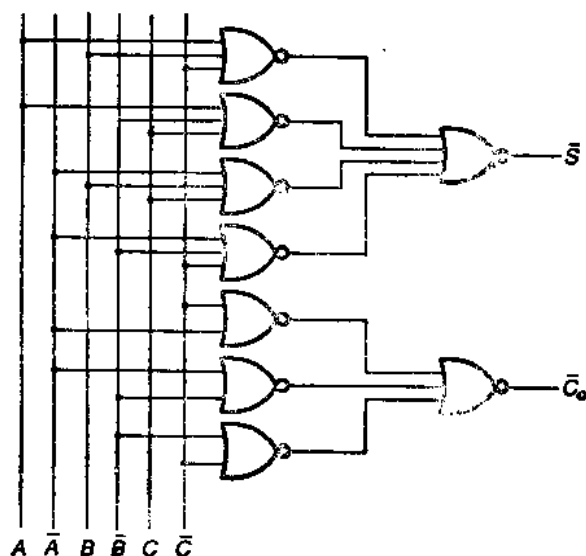


图 1.19 一个用或非门实现的全加器

其中 A 、 B 、 C 为输入， S 为和输出， C_o 为进位输出。与非门与或非门类似，只是下拉器件是串联的，这样，只有当所有下拉晶体管都导通时输出才被拉到低电平。注意下拉晶体管增加时，下拉通路的电阻也增加。为了使输出仍能被拉到低电平（小于 V_T ），下拉晶体管的形状因子对上拉晶体管之比必须增加，通常这就要减小上拉晶体管的形状因子或是增加下拉晶体管的形状因子。图 1.20 所示为一个 3 输入端与非门。

可以有更复杂的逻辑门。图 1.21 所示为一个复杂的门结构，它包含几个串并联的

1) 注意，若有其它供应电流的源，输出完全可以升到 V_{DD} 甚至更高。当它升到高于 V_{DD} 时，源变成了漏，而漏变成了源。这样配置的晶体管称之为“二极管”接法。



图1.20 一个3输入端或非门

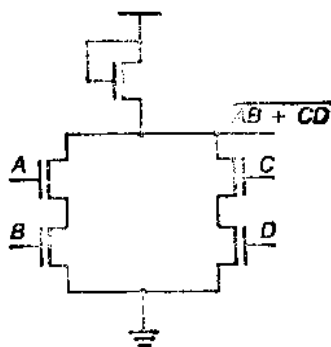


图1.21 一个复杂门

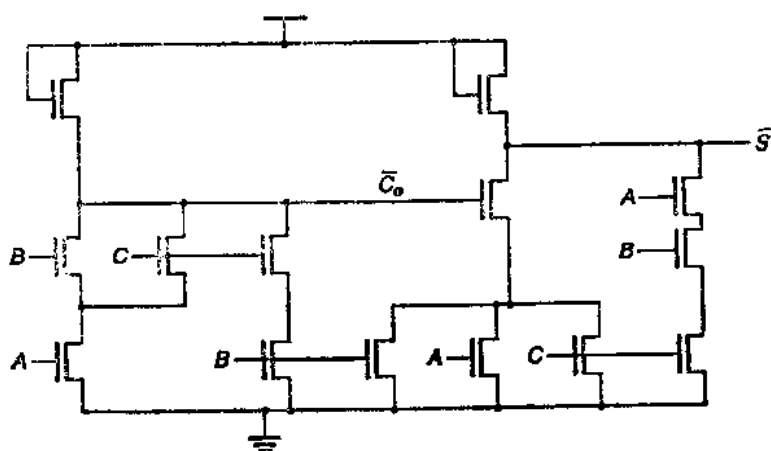


图1.22 用复杂门实现的一个全加器

下拉晶体管。至此所讨论的所有的初级门都是反相的。图 1.22 所示为一个简化的 常用全加器逻辑，它采用了复杂门[Carr 72, P.134]。

图1.16(b)所示的选择器是最简单的传输晶体管逻辑网络之一。选择器工作原理如下：当 S 为高电平时，输入 I_1 连接到输出 Z ；当 S 为低电平时，输入 I_0 连接到 Z 。即使 I_1 和 S 都为 V_{DD} ，输出也只能升到 $V_{DD} - V_T$ 。但如果 I_1 和 I_0 为 $V_{DD} - V_T$ ，输出也仍能升到 $V_{DD} - V_T$ 。如果 S 和 $\bar{S}$ 在某种情况下同时为高电平，这将导致 I_1 与 I_0 连接(潜通道)，这是不需要的。图 1.23 所示为一个用传输晶体管实现的全加器。

我们已经考察了两种形式的 MOS 数字电路（传输门和反相逻辑门）的特点。我们知道，可以用这些电路形式来组成任何大型的数字功能块，这是因为任何一种形式都包括二输入端或非门，而或非门本身是普遍适用的。下一步我们就来考察这些电路部件如何组合成一个相容的合理的电路家族，以构成各种组合逻辑。为此，我们需要扩大（有些情况下需要改变）这些部件的组成规则。我们需要有关这些部件连线的附加规则。确定电路的组成规则，很少采用严谨的方式¹⁾，不过，了解电路如何构成，在何种条件和

1) 如果是采用严谨方式的话，那就太好了，因为这样我们就可以证明组成的电路是否正确。不幸的是，这是个很困难的问题。

限制下它们可以连在一起，是 VLSI 设计的关键之一。这些组成规则是电路设计方法的一个要素。

让我们从全局的角度重新仔细考察一下增强型负载逻辑门。构成这种逻辑门要有几条组成规则。逻辑门包含一个上拉的部分和一个下拉的部分。上拉的部分由一个增强型

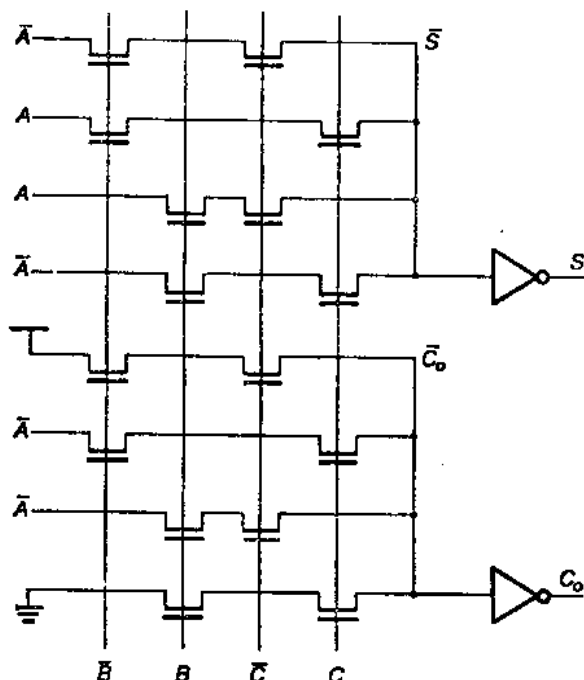


图 1.23 一个用传输晶体管实现的全加器

晶体管构成，其漏与栅接 V_{DD} ，源接输出。下拉的网络包含一些晶体管，它们的源和漏串并联成一个二端网络，一个端点接地，另一端连输出节点。这些下拉晶体管的栅就是逻辑门的输入端。输入电压为 $V_{DD} - V_T$ 时代表高电平，为 V_T 或更低时代表低电平。如果导通的晶体管没有接地通路，则输出被拉到高电平。如果输出端和接地端之间接有导通的晶体管又假定导通器件的栅电压为 $V_{DD} - V_T$ ，则晶体管的形状因子必须设计得使其在每一种可能的通路下都可以把输出端拉到低于 V_T 。换句话说，有效的输入高电平产生有效的低电平¹⁾，反之亦然。我们把这些逻辑门称之为“I 型”逻辑门。

I 型逻辑门对其它 I 型逻辑门没有扇出限制（这里不考察动态特性），因为 I 型逻辑门的直流输入阻抗为无穷大。另一方面，把两个 I 型逻辑门的输出端接在一起则会引起操作错误，因为这时两个上拉器件是并联在一起了。一个下拉通路的导电能力足以把一个上拉器件拉低，但可能不足以把两个上拉器件拉低。

由此我们有第一个组成规则：不允许把两个 I 型逻辑门的输出端连在一起。这一限制使系统设计容易一些。注意，违反这一组成规则的电路可能不工作，但并不是说它一定不工作。有人曾经设计过一些有两个上拉器件的逻辑门，它们确实工作得很好。不过，我们现在是利用这第一个组成规则，用最简单的方法，排除一些逻辑门，以简化电路设计工作的复杂性。用这种方法来构造逻辑门，不但会使设计容易一些，也使查错容易一些。

在增强型传输晶体管逻辑中，存在两个难题，即避免潜在通道及跟踪由于阈值而产生的压降。假定传输晶体管栅和漏分别加高电平电压 v_G 和 v_D ，则其源电压 v_S 在 v_D 最小值和 $v_G - V_T$ 之间变化。若 v_G 比 V_T 小，则输出与 v_D 无关。图 1.24 所示为一个具有各种不同阈值压降的传输晶体管网络，阈值压降的值标在电路图上。

当传输晶体管与增强型负载逻辑门组合在一起时，我们发现存在矛盾。增强型负载

1) 第四章将进一步加以扩展，把噪声容限也考虑在内。

反相器的最高输出电压为 $V_{DD} - V_T$ 。如果这一电压再传到传输晶体管的栅上，则从传输器件输出的最高电压为 $V_{DD} - 2V_T$ 。对于第二个增强型负载逻辑门来说，这一电压作为高电平就不够了。更坏的是，如果允许把传输晶体管任意连接起来，输出电压可能低至 $V_{DD} - 3V_T$ ， $V_{DD} - 4V_T$ 甚至更低。

可以设想有几种解决方法。最简单的办法是把传输晶体管排除在外，有时就是这样做的。另外的解决办法是设计一种类型的增强型负载逻辑门，它的形状因子调整成即使栅电压仅为 $V_{DD} - 2V_T$ ，输出也可被拉到 V_T 以下，这种逻辑门称为“Ⅱ型门”。若下拉晶体管尺寸相同，则上拉器件形状因子就要比Ⅰ型门小。但是仍然存在传输晶体管输出电压为 $V_{DD} - 3V_T$ 以至更低的问题，这可以通过在设计方法上再加一条限制来避免：不允许传输晶体管网络的输出接到其它传输晶体管的栅上。图 1.25 所示的逻辑电路就包括了Ⅰ型和Ⅱ型增强型负载逻辑门和传输晶体管网络，其高电平电压不低于 $V_{DD} - 2V_T$ 。表 1.2 列出了何种类型的逻辑门可以接受何种输入信号。

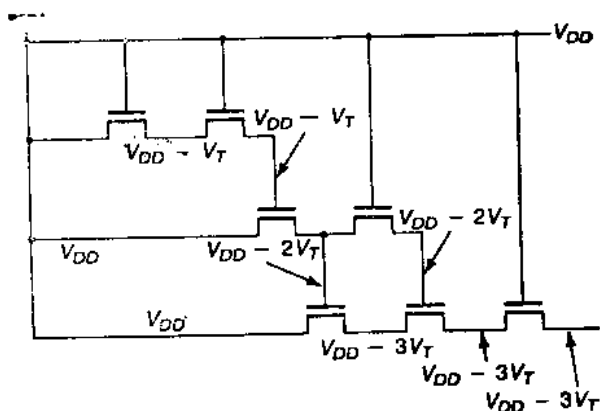


图 1.24 一个增强型 n 沟传输晶体管的网络，图上标明了阈值压降。假定每一节点对地都有大的漏电阻

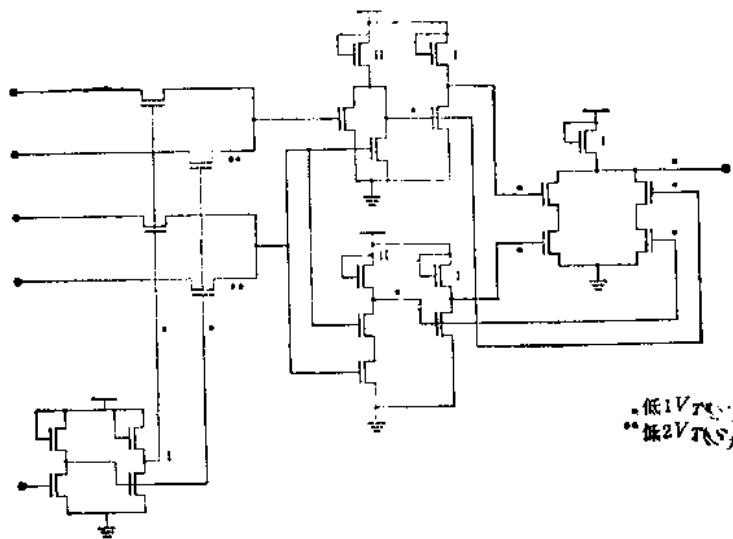


图 1.25 采用一种简单方法的阈值压降

表 1.2 设计方法规定

	Ⅰ 型	Ⅱ 型
从逻辑门输入	可以	可以
从传输晶体管输入	不可以	可以
输入 $V_{DD} - 3V_T$	不可以	不可以

我们已成功地研究出了一种相当初步的全增强型数字 MOS 组合电路的设计方法。这种方法使用了三种基本的电路形式。现代的 VLSI 可能有五十种这样的电路形式。尽管 50 比 3 大得多,但它仍比 VLSI 芯片中的门数小几个数量级。这种电路设计方法可以应用于许多芯片。观察发现,电路设计方法的复杂性随所设计的系统规模的缩小而缓慢减小。有两个因素使设计方法变得比较复杂。第一个因素是,在很大的系统中会有一些新的现象,例如金属的电阻变得重要了。这就可能需要更复杂的设计方法来对付更复杂的情况。第二个因素是对性能的追求。比较复杂的方法能够得到较高的性能,尽管这不一定是设计电路所必需的。就以我们刚讨论的电路为例,我们可以把所有 I 型逻辑门换为 II 型逻辑门,这样就使方法简化了。这可以使电路形式的数目减少 30%,但由于上拉电阻较大,这也使电路的速度变慢了。从历史上看,在基本单元的简单化与其性能的折衷方面,工程界与计算机科学界曾倾向于简单及可设计性,但这仅仅限于代价不大的情形。关于这些因素的充分讨论不属本书的范围。不过应该指出的是,技术发展的趋势是这样的:按过时技术设计的产品是落后于时代的。

自举逻辑门

增强型负载逻辑门的输出不会升到 $V_{DD}-V_T$ 以上,所以可以说在 V_{DD} 与 $V_{DD}-V_T$ 之间的这部分电源电压被浪费掉了。在第四章中我们还将看到它的速度也比较低,功耗-延迟积较差。增强型上拉器件不能把输出拉到 $V_{DD}-V_T$ 以上的原因是它的栅电压固定为 V_{DD} 。避免这一问题的一种简单办法是把所有上拉晶体管的栅都接到第二个电源电压 V_{oc} 上 (V_{oc} 比 $V_{DD}+V_T$ 大),其代价是要增加一个电源并增加把这一电源分配到整个芯片所需的布线,不过新增电源消耗的电流很小,而且其逻辑门输出的高电压可增加到 V_{DD} 。

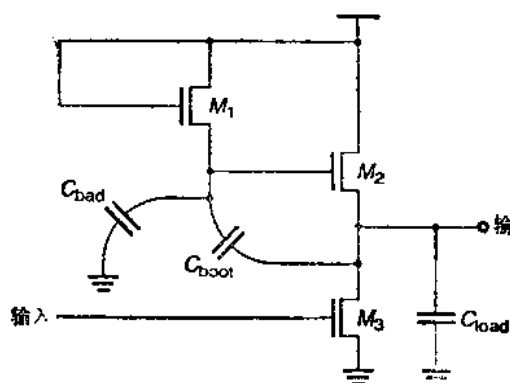


图 1.26 自举反相器

另外一种技术是把上拉晶体管的栅电压¹⁾举到电源电压以上 [Joyhson 72]。图 1.26 示出了一个自举反相器。当逻辑门的输入为高电平时,输出节点被下拉到 V_{OL} 。 V_{OL} 定义为输出逻辑低电压, V_{OL} 小于 V_T 。晶体管 M_1 的源为 $V_{DD}-V_T$ 。在电容 C_{boot} 上的电压为 $V_{DD}-V_T-V_{OL}$, 电荷为 $(V_{DD}-V_T-V_{OL})C_{boot}$ 。 M_1 源的寄生电容 C_{bad} 上的电荷为 $(V_{DD}-V_T)C_{bad}$ 。当输入电压向 V_T 下降、输出电压从 V_{OL} 上升时,

M_1 向电容 C_{load} 充电。由于 M_1 不通,在自举节点上的电荷被陷住。总电荷 Q_{boot} 为:

$$Q_{boot} = (V_{DD}-V_T-V_{OL})C_{boot} + (V_{DD}-V_T)C_{bad} \quad (1.6)$$

电容 C_{boot} 将力图使其上的电压保持为常数。这样,当 v_{out} 上升时, v_{boot} 也上升。为了使 v_{out} 达到 V_{DD} , v_{boot} 必须达到 $V_{DD}+V_T$ 。我们可以推导出 C_{boot} 为此必须满足的条件。当 v_{boot} 达到 $V_{DD}+V_T$ 时, C_{bad} 上的电荷为 $(V_{DD}+V_T)C_{bad}$, C_{boot} 上的电荷为 $V_T C_{boot}$ 。新的在

1) 英文原意为“用鞋带把自己拉起来”。

自举节点上的电荷为二者之和，它必须仍然等于 Q_{boot} 。由此有

$$Q_{boot} = C_{bad}(V_{DD} + V_T) + C_{boot}V_T \quad (1.7)$$

由此可解出为保证输出升到 v_{DD} 所需的 C_{boot} 最小值：

$$C_{boot} = \frac{2V_T}{V_{DD} - 2V_T - V_{OL}} C_{bad} \quad (1.8)$$

C_{boot} 可以简单地用一个晶体管来做成，如图 1.27 所示。把晶体管的源和漏短接在一起即可。往往晶体管的本征栅电容就够了。 C_{bad} 来源于布线与晶体管的电容。注意，在 $C_{bad} = 0$ 的极端情形， C_{boot} 两端的电压为常数。

自举是一种极为有用的技术，并以多种形式为本书所普遍采用。它使我们可以取消 V_{GG} 电源及相应连线，代价仅是增加一个晶体管。在 VLSI 中用晶体管替换引线往往是合算的。

自举负载有些限制，这些限制有些时候要影响它的使用。一个限制是时间。由于 MOS 晶体管在不导通时的电阻并不是无穷大，经过若干毫秒，自举节点上的电荷就会漏掉，输出将漂回 $V_{DD} - 2V_T$ 。若芯片必须周期性地停止工作，或芯片的一部分有可能长期无动作，就不宜采用自举。

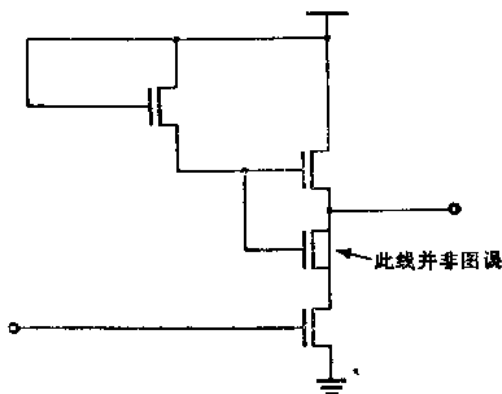


图1.27 用MOS电容的自举反相器

耗尽负载

自举负载把输出一直拉到 V_{DD} 是因为其 v_{os} 电压近似保持不变。回顾 1.1 节可看出，耗尽型晶体管可以被设想为一个在栅与源之间有一个自建电池的增强型器件。要做一个

耗尽型负载，我们只需简单地把耗尽型晶体管的栅与源连在一起 [Masuhara 72]。其代价是制造工艺稍稍复杂些，成本稍高些。图 1.28 所示为一对耗尽负载反相器的版图。请注意，在耗尽型上拉晶体管上，作为一种耗尽型阈值调整版的速记符号，我们往往在耗尽型晶体管的栅上加一星号(*)。* 字号(#)则用以标志零阈值器件。如果图 1.25 中所示电路重新通为耗尽型门，则最低的高电平电压将从 $V_{DD} - 2V_T$ 增至 $V_{DD} - V_T$ 。

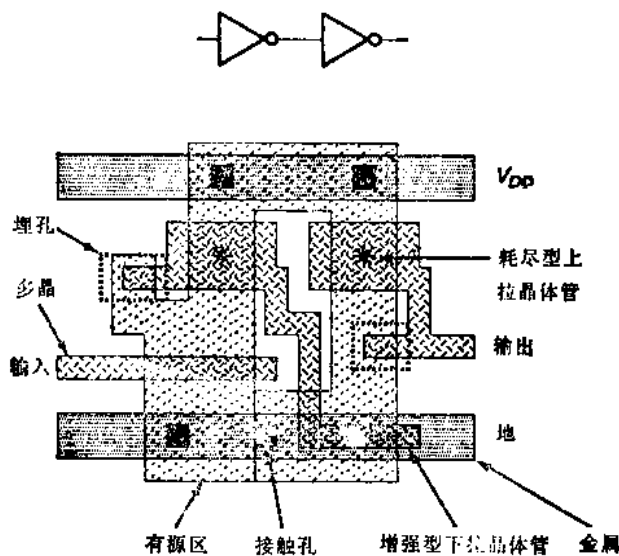


图1.28 一对耗尽型负载反相器的版图

图 1.29 所示为串接的两个

反相器的瞬态响应。注意，上拉响应比下拉响应慢得多。这是因为下拉晶体管导电必须比上拉器件好，这样，当输入为高电平时才能把输出拉到 V_T 以下。对下拉静态特性的这一限制转化成为动态响应的非对称性。

驱动某固定容性负载的逻辑门的上拉速度，总是可以通过增加该逻辑门中所有晶体管的宽度来提高的。但这样做有两个问题，第一个问题是有比逻辑电路的功耗也增加了¹⁾；第二个问题是输入电容（它也是前级逻辑门的负载电容）也增加了。通过巧妙的电路设计，人们可以在一定限度内绕开这些问题。

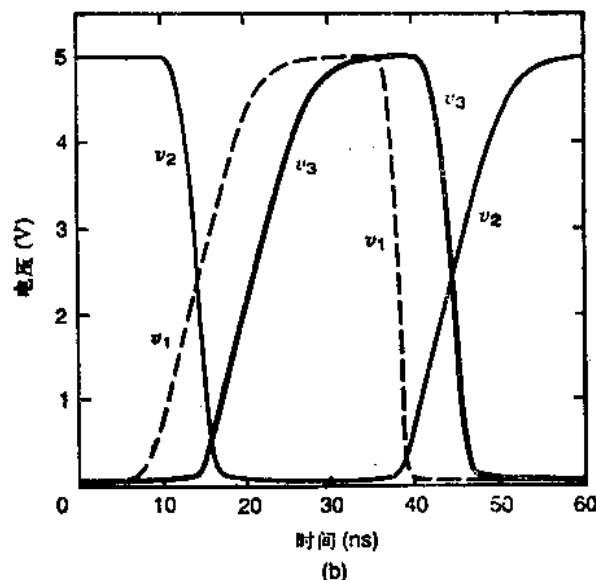
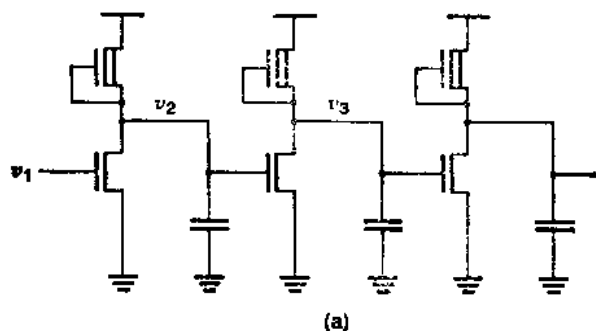


图1.29 一串耗尽负载反相器的瞬态响应
(a) 电路图；(b) 波形。注意上升时间慢

管的宽度来提高的。但这样做有两个问题，第一个问题是有比逻辑电路的功耗也增加了¹⁾；第二个问题是输入电容（它也是前级逻辑门的负载电容）也增加了。通过巧妙的电路设计，人们可以在一定限度内绕开这些问题。

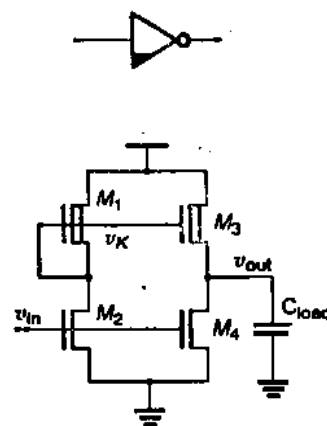


图1.30 一个反相超级缓冲器

一种加速大负载驱动的办法是采用超级缓冲器(Mead 80, pp. 17—18)。图1.30画出了一个反相超级缓冲器。这一部件用了4个晶体管而不是两个。若晶体管的尺寸适当，则超级缓冲器可以提高耗尽负载反相器驱动大电容的速度，而不增加静态功耗。超级缓冲器的工作机理如下：设输入为低，输出及内部节点电压 v_K 均为 V_{DD} 。输入的上升将使两个下拉晶体管都导通。内部节点电压 v_K 将首先下降，这是因为它的负载（由 M_3 的栅和某些寄生电容组成）比 C_{load} 小。这使得由 M_3 流出的电流比 M_3 的栅接源时流出的电流稍小一些。这一效应并不很强，超级反相器的下拉速度与常规反相器差不多，这是因为放电动态特性主要是由下拉晶体管的电导决定的。这一点问题不大，因为下拉速度本来就较快。对上拉的情形，反相器的输入由 V_{DD} 变为 V_{OL} 这使两个下拉器件都断开。起

1) 功耗正比于电流乘以电源电压 V_{DD} 。

初两个上拉晶体管的 v_{GS} 均为低, 随着 v_k 迅速上升到 V_{DD} , 当负载电容很大时, M_3 的栅对源电压接近于 $V_{DD} - V_{OL}$ 。由于这一上冲, M_3 在输入上升的最初若干毫微秒内导通很强。这可以使得上拉速度比简单的反相器在相同功耗及相同负载电容的情形下快 2 到 6 倍。从我们对其动态特性的描述中应该可以看出, 只有当驱动大的电容时才有这种好处。两种 nMOS 超级缓冲器的版图示于图 1.31 和图 1.32。

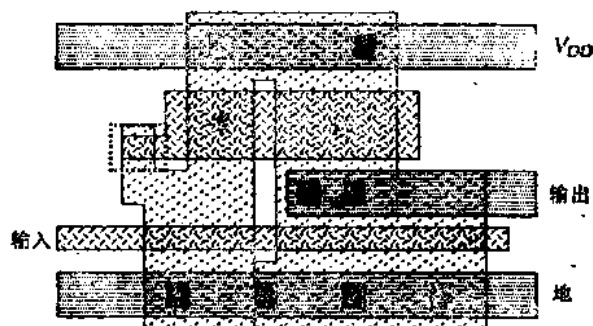


图 1.31 一个 nMOS 超级缓冲器的版图, 它有一个上拉输出的耗尽跟随器

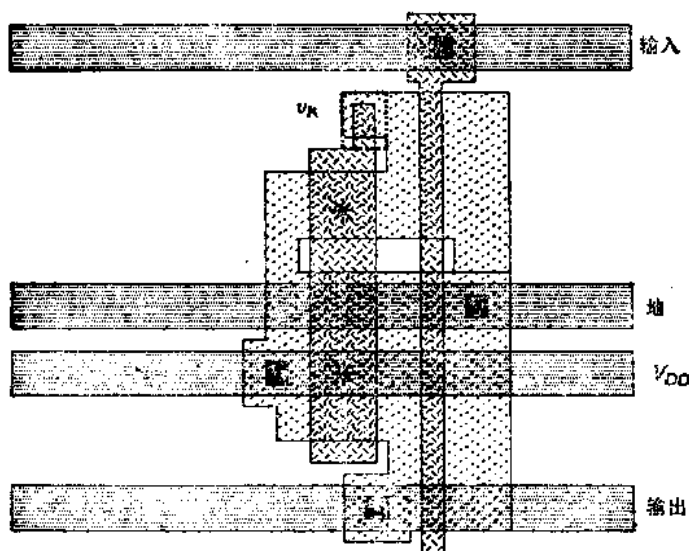


图 1.32 与图 1.31 电路相同的超级缓冲器的另一种版图

超级缓冲器说明了在设计 VLSI 时常常采用的一种思想模式。在超级缓冲器中, 我们可以部分地把动态与静态特性问题分开, 从而达到效果较好的设计。这是以电路的复杂性为代价的¹⁾。在电路设计者的窍门中包含有许多技巧, 超级缓冲器仅仅是头一个。

在许多 nMOS 工艺中, 除了增强型及耗尽型器件之外还有一种“零”阈值晶体管。这种器件对缓冲器及小扇入的选择器是极为有用的。图 1.33 所示为采用了零阈值器件的缓冲器。与以尺寸相同的耗尽型晶体管作负载的超级缓冲器相比, 零阈值缓冲器

1) 与此相对, 从自举增强型负载过渡到耗尽型负载, 是以工艺的复杂性为代价的, 而电路复杂性却是下降的。

功耗小得多，但速度也降低不少，而且也输出不了那么高的电压（最大输出电压为 $V_{DD} - V_{Tn}$ ）。但另一方面，我们可以加宽零阈值上拉器件而不增加静态功耗，这是因为它仅仅在开关过程中导通。对于较大的负载，零阈值缓冲器的速度可接近于超级缓冲器，而功耗仍较小。

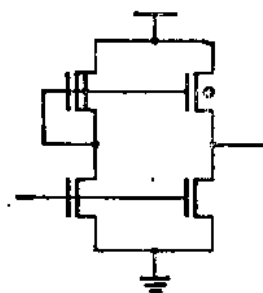


图1.33 输出级上拉器件阈值为零的超级缓冲器。与耗尽跟随超级缓冲器相比，功耗较低速度较慢

零阈值晶体管在 $v_{GS} = 0$ 时是有漏电的。因此我们必须注意不要让很多个“关断”的零阈值晶体管影响导通晶体管的状态。这就是零阈值器件仅仅适用于小扇入的选择器的原因。对于大扇入的情形，例如母线，就必须用增强型晶体管。

零阈值器件使设计方法进一步复杂化。如果耗尽型晶体管是现成的，通常我们就把自举增强型负载换为耗尽型负载。但即使没有零阈值器件，仍然需要考虑两种有比的逻辑门。这就是其输入为其它耗尽型负载逻辑门输出（可上升到 V_{DD} ）的逻辑门和输入来自传输晶体管的逻辑门。有了零阈值器件，我们就需要考虑第三种有比逻辑门（三种门分别对付输入为 V_{DD} ， $V_{DD} - V_{Tn}$ ，及 $V_{DD} - V_{Tz}$ 的情形）或者归并为两类，例如，我们可以把一种逻辑门用于输入为 V_{DD} 以及比 V_{DD} 低一个零阈值电压的情形，而把另一种逻辑门用于输入比 V_{DD} 低一个增强型阈电压的情形。

CMOS静态逻辑形式

至今我们所讨论的所有逻辑门在输出为低时都要消耗直流功耗。这一热量要传出芯片是很费力的，而MOS晶体管的性能又随芯片温度的上升而下降，这就使得nMOS VLSI电路的设计变得十分复杂。为了降低功耗必须采用特殊的电路形式。几种这样的特殊nMOS电路形式将在下一节中讨论。

一般说，采用比较复杂而费钱的制造技术可以使电路设计工作好办一些。上面我们已经介绍了三种nMOS晶体管。它们的共同特征是晶体管栅电压上升，晶体管的电导就增加。而p沟晶体管的工作正相反，当栅电压上升时，pMOS晶体管导电性减弱。这一点有利于建立推挽式结构，它仅仅在开关时才消耗功率。

图1.34所示为CMOS反相器，或非门及与非门（Wanlass 63）。注意其下拉网络与nMOS有比逻辑的情形完全一样。上拉网络与下拉网络一一对应。在下拉网络中有两个串联器件，与此相对应，在上拉网络中就有两个并联器件，在下拉网络中有两个并联器

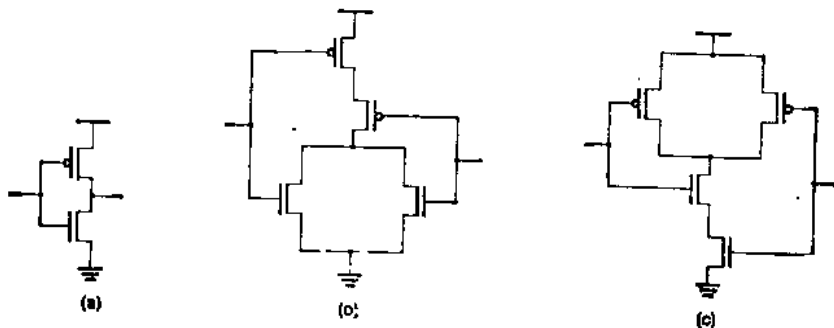


图1.34 CMOS逻辑门电路图，(a) 反相器，(b) 或非门，(c) 与非门

件，在上拉网络中与它们相对应就有两个串联器件。全部上拉晶体管都是增强型P沟器件，而全部下拉晶体管都是增强型N沟器件。在设计得当的这类CMOS逻辑门中，当上拉网络导通时，下拉网络就关断，反之亦然。只有在开关的瞬态过程中，上拉和下拉网络才可能同时导通。

CMOS反相器的工作机理如下：当输入为高电平（ V_{DD} ）时，下拉晶体管导通，但P沟器件的 v_{GS} 为零，因而上拉晶体管关断。因此，输出被一直拉到地。当输入为低电平（0伏）时，P沟器件的 v_{GS} 为 $-V_{DD}$ ，因而它导通，下拉器件则是关断的。输出一直升到 V_{DD} 。

上拉与下拉网络的对偶特点可以借助于该门所完成的布尔函数来理解。以描述全加器进位输出 C_o 的布尔函数为例。设 Z 为 C_o 的反相，则有

$$\bar{Z} = AB + BC + AC \quad (1.9)$$

其中

$$Z = \bar{C}_o \quad (1.10)$$

下拉网络可以直接从方程（1.9）读出来，只要我们把“与”解释为串联而把“或”解释为并联。利用德摩根定理，有

$$Z = (\bar{A} + \bar{B})(\bar{B} + \bar{C})(\bar{A} + \bar{C}) \quad (1.11)$$

由这一方程可以直接读出上拉网络结构。方程（1.9）与（1.11）是互相对偶的，每个“或”翻为“与”，每个“与”翻成“或”，每个符号都被反相。

方程可以被进一步简化，而且这些简化的方程可进一步翻成图的形式。我们有

$$\bar{Z} = A(B + C) + BC \quad (1.12)$$

$$Z = \bar{A}(\bar{B} + \bar{C}) + \bar{B}\bar{C} \quad (1.13)$$

方程（1.9）至（1.13）所对应的电路图示于图1.35（注意方程（1.12）与方程（1.13）并无前面的对应关系）。

在CMOS中我们可以采用传输晶体管电路而不会招致任何阈值压降。这是通过采用并联的P型及N型传输晶体管达到的，每当传输门既需要传输高电平也需要传输低电平时我们就这样做。图1.36为用CMOS完成的一个四输入选择器，一个输入端接到 V_{DD} ，还有一个接到地。对于这样的电路分支就可能采用全N型（只传输低电平）或全P型（只

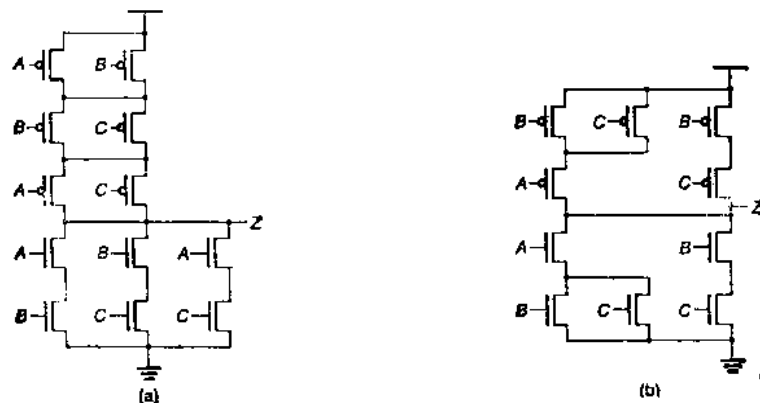


图1.35 (a) 全加器进位输出的CMOS电路图；(b) (a) 中电路的优化形式，它用的晶体管较少，上拉的电阻较低

传输高电平) 晶体管。

许多电路把传输晶体管与恢复逻辑门协调地用在一起。图1.37所示为异或门(XOR)及异或非(XNOR)门的几种电路方案。

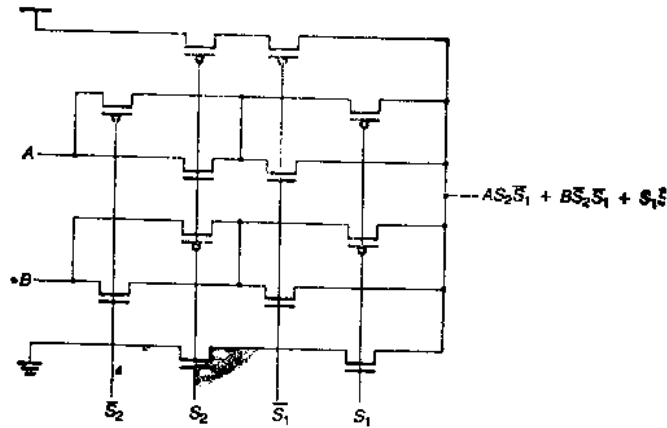


图1.36 带有两个输入为常数的四对一选择器

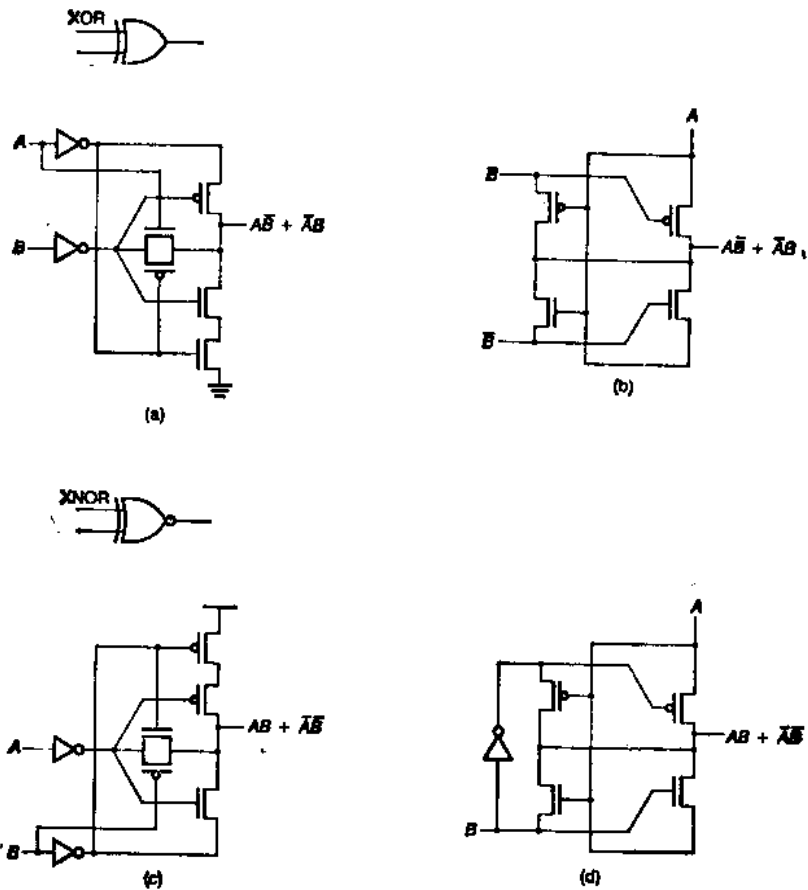


图1.37 异或及异或非电路。(a)及(b)为异或电路及其等效电路,(c)及(d)为异或非及其等效电路。这些电路把恢复逻辑门与传输晶体管逻辑结合起来了

例1.2 对偶性。

若 $F = AB + \bar{C} + D$, 则 $\bar{F} = (\bar{A} + \bar{B})C\bar{D}$ 。

例1.3 几种类型反相器的动态响应。

图1.38至1.42示出了几种不同类型的反相器的动态响应, 其中包括一个增强型负载反相器, 一个自举反相器, 一个耗尽型负载反相器, 一个耗尽跟随超级缓冲器, 和一个经典的CMOS反相器。

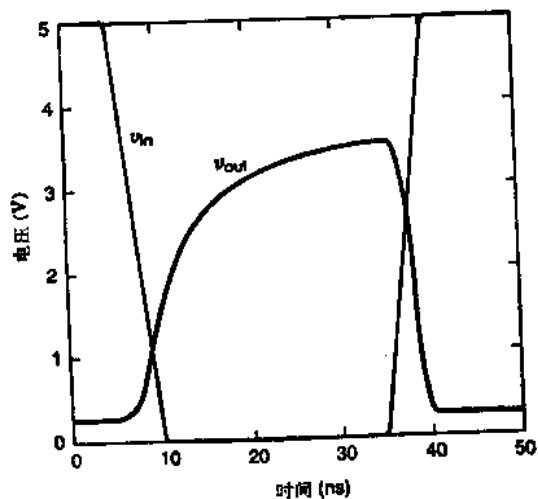
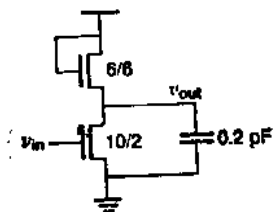


图1.38 一个全增强型反相器的电路图及波形图

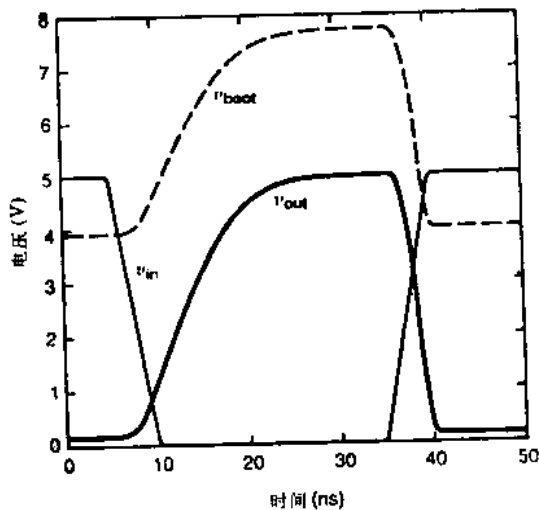
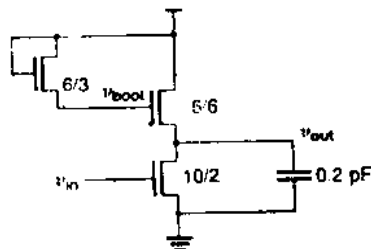


图1.39 一个全增强型自举反相器的电路图及波形图

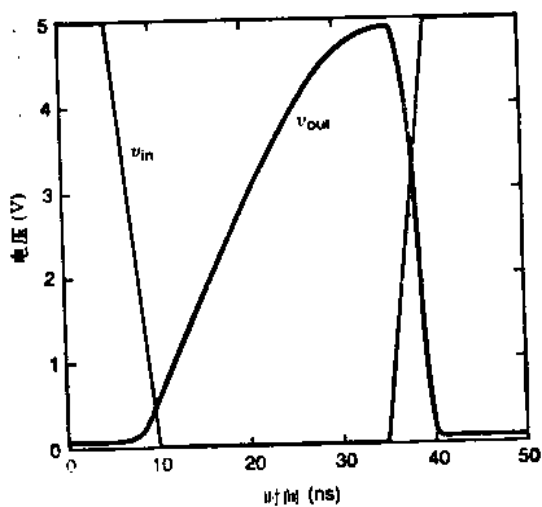
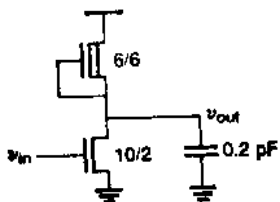


图1.40 一个耗尽负载反相器的电路图及波形图

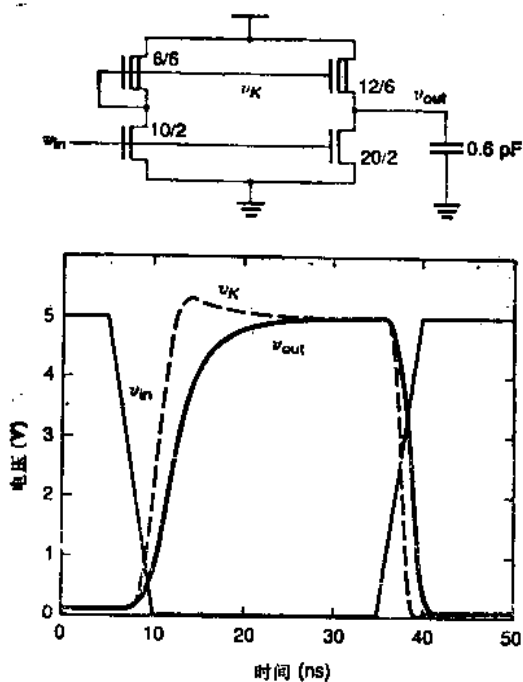


图1.41 一个耗尽型跟随超级缓冲器的电路图与波形图

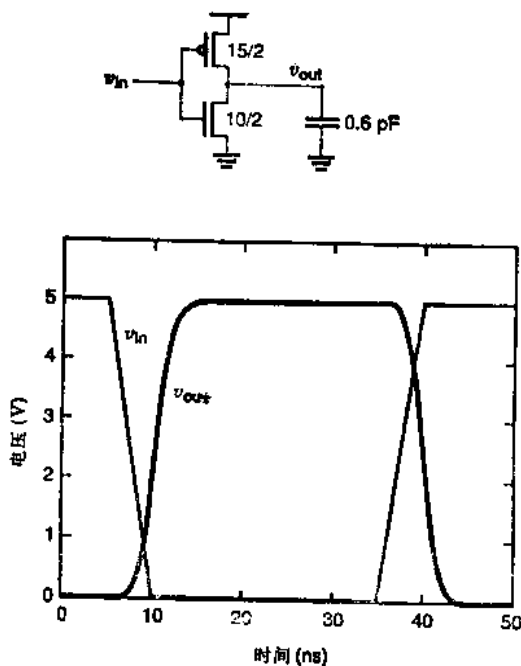


图1.42 一个CMOS反相器的电路图与波形图

1.3 存贮元件及顺序电路

在1.2节中我们考察了用于实现组合逻辑的各种电路形式。在组合逻辑电路中，输出仅仅为当时输入的函数。组合逻辑是建造大型数字系统所必需的，当它与存贮器结合起来时，会变得极为有用。一个典型的数字系统必须经过一系列操作，而每一个操作的结果均取决于其先前操作的结果。对于这样的顺序系统，其输出取决于输入的历史。

顺序系统是由存贮元件和组合逻辑构成 [Hill 74]。MOS 电路有几种存贮元件。静态存贮元件是由反馈形式的逻辑门构成的。图1.43所示为由交叉耦合的或非门组成的简单 SR 触发器的电路图。图 1.44 和图1.45为其两种版图。锁存器当 S 和 R 均为低时有两个稳态。

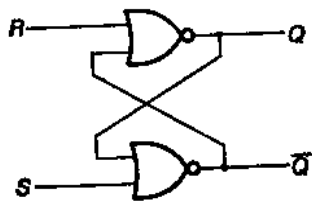


图1.43 交叉耦合的或非门

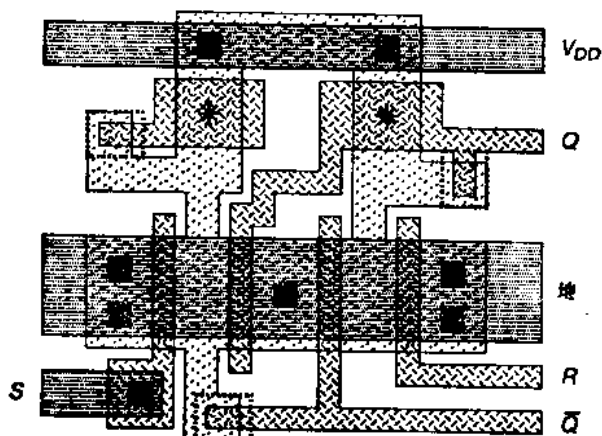


图1.44 图1.43所示电路的版图

除了静态存贮元件之外，MOS技术还提供了另一种选择，即动态存贮元件。静态存

贮器只要电源不断就可以“记住”它的状态¹⁾，而动态存贮只能把它的状态记住一段够用的时间，但不能永远记住。动态存贮元件的好处是它可以比静态存贮元件做得密集得多。在MOS技术中，动态存贮单元可维持其信息约1毫秒，与温度有关。因此，一个单元存贮的周期比一级反相器的最小延迟可以长 10^6 倍。从原理上说，MOS动态存储单元包括一个开关和一个电容。一个很好的类比是，恢复逻辑与静态存贮可以积极地抵消噪声效应，而传输晶体管与动态存贮器则不能。

一个顺序系统可以由许多种方法来实现。最普通和直接的方法是采用一个中央时钟来同步顺序操作。中央时钟给全芯片提供了统一时刻表，有利于数据在芯片上有次序的运动。

时钟周期通常分为若干个子周期，称之为相，它提供了更细微的时间分度。时钟不但可用于同步，而且其波形也可用于电路范畴。时钟也是一个很好低阻抗动态电源。

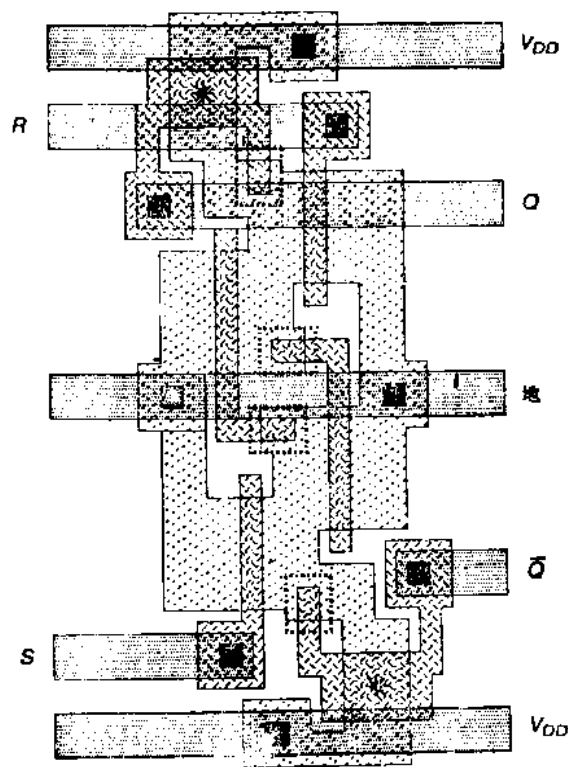


图1.45 图1.43所示电路的另一种版图

移位寄存器与锁存器

移位寄存器是研究时钟电路作用的一个极好例子。原理示于图1.46的移位寄存器是利用在任何逻辑设计的基础书中都可以找到的经典电路形式来实现的。实际上，TTL [TJ 76] 设计者所熟悉的全部钟控和非钟控的锁存器、寄存器，及触发器家族都可以很容易地用MOS技术来完成，这样的电路的一个例子示于图1.47。

图1.48所示为采用动态技术，用CMOS及耗尽负载nMOS来实现的图1.46所示移位

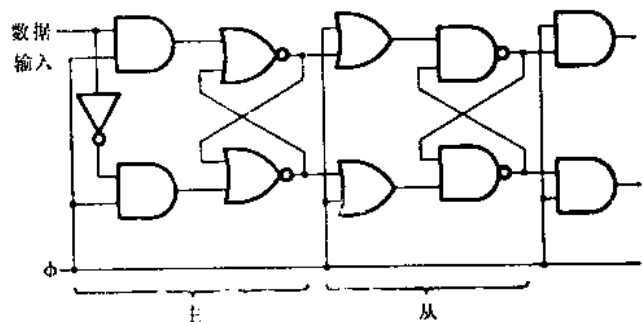


图1.46 一个主-从移位寄存器

1) 在磁泡存储器及EPROM中，即使断电状态也可被“记住”，断电后仍保持记忆的存储器称为非挥发性的。本书中所讨论的所有MOS存储器都是挥发的。

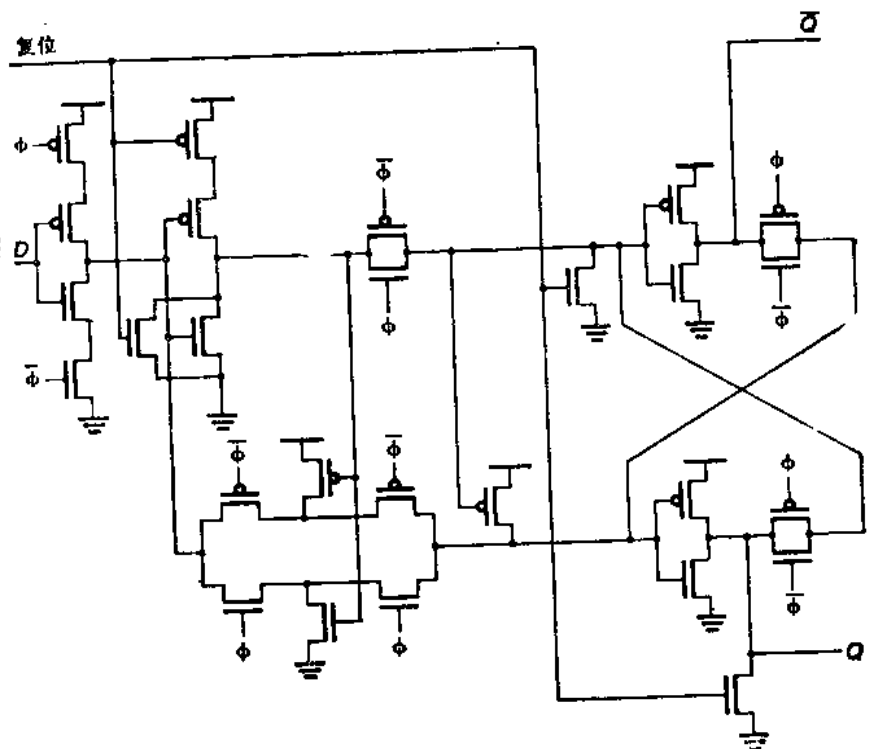


图1.47 一个带复位的 CMOS D 触发器

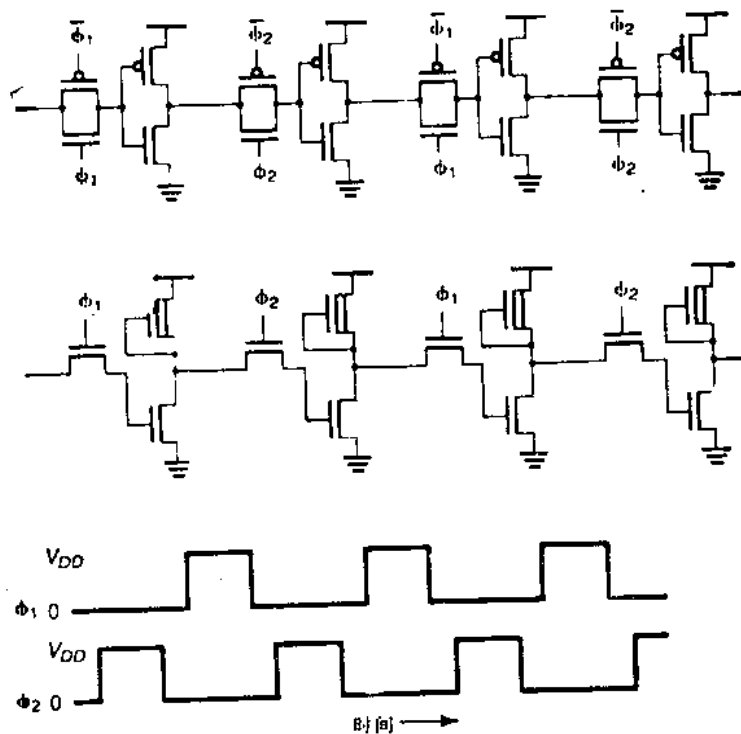


图1.48 用CMOS及耗尽型负载nMOS实现的动态移位寄存器。图中还画出了互不重叠的时钟波形

寄存器.数据是存在反相晶体管栅电容及布线电容上的。传输晶体管的作用象一个开关,当导通时允许电容上的电荷流出流进,而关断时则把电荷陷住。时钟分成互不重叠的两个相。在这种方法中(Mead 80, p.65),时钟波形 ϕ_1 和 ϕ_2 不能同时为高,否则数据就会从移位寄存器的一端走到它的另一端。这可以与运河中禁止同时打开所有船闸相类比。我们在讨论时钟电路时,往往希望用统一的方法处理CMOS和nMOS电路。因此我们引入了一个超越具体实现技术的传输门符号,就像与具体技术无关的标准反相器符号那样。图1.49所示为常开及常闭*传输门及实现该门的MOS电路。图1.50所示为用这种通用符号画出的图1.48所示的动态移位寄存器。另一种有用的符号示于图1.51。它代表源接地的一个下拉晶体管。我们还用这种新符号画出了一个或非门。这一符号在阵列中最有用。

图1.52画出了另一种普通CMOS移位寄存器电路。其中,时钟控制的晶体管与反相器晶体管的源串联。这种电路与图1.48中用传输晶体管的方案相比在版图方面有好处。

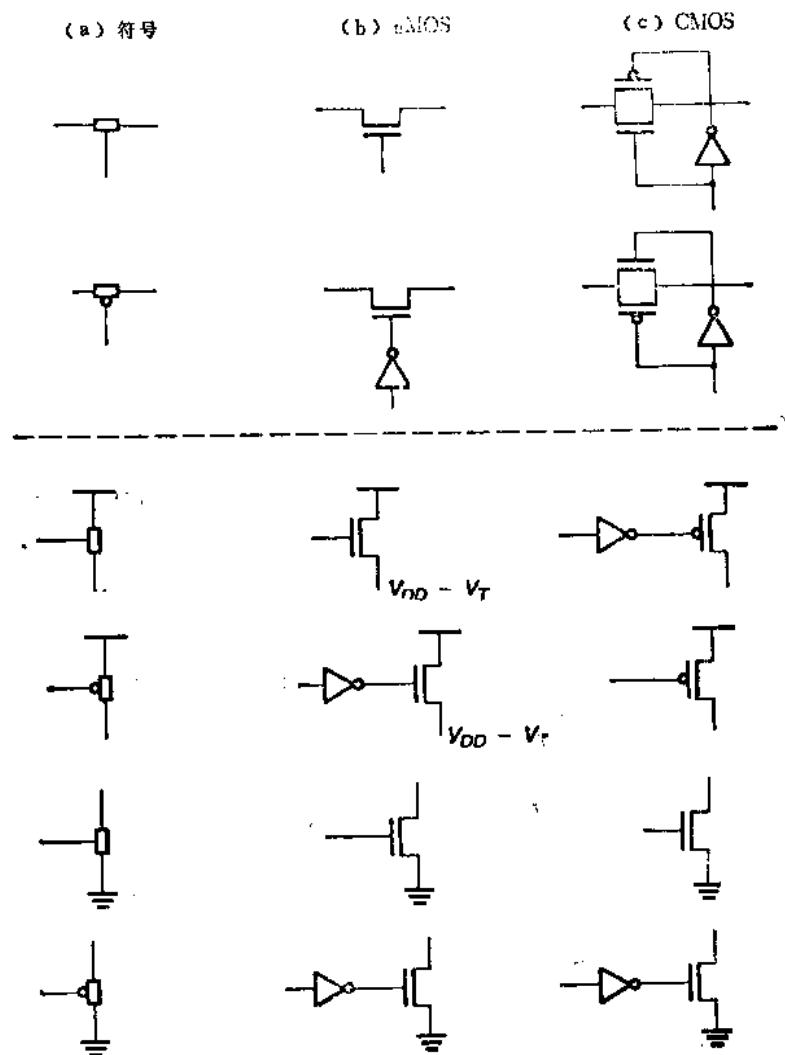


图1.49 传输门的一般符号

*常开及常闭均指输入为高电平时的状态。——译者注

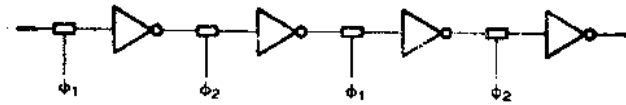


图1.50 MOS 移位寄存器的一般符号

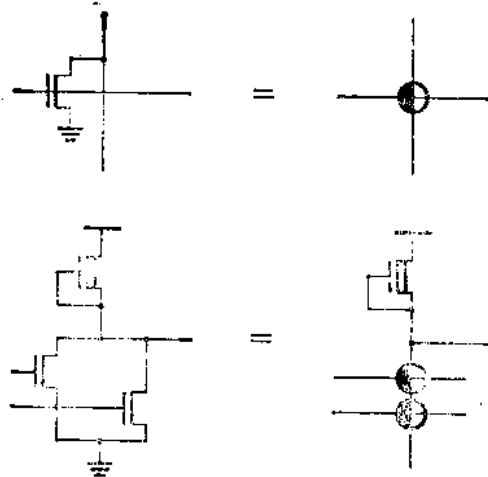


图1.51 下拉符号，对阵列特别有用

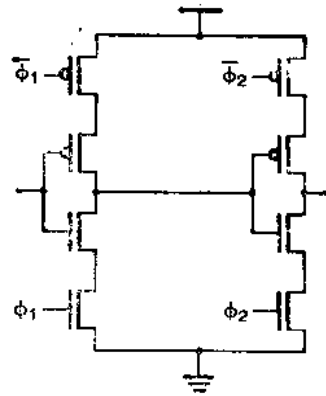


图1.52 另一种 CMOS 移位寄存器单元



图1.53 采用图1.48方案的CMOS移位寄存器显微照片

这些移位寄存器的显微照片示于图1.53和图1.54。

随着电路的每一种革新，都会带来新的失误的机会。钟控电路技术引入了“电荷分配”的毛病。当两个电压不同的电容通过传输晶体管连在一起时就会发生电荷分配的问题。当传输器件导通时，两个电容上的电压会平衡到某个中间值。我们取一个有点人为的移位寄存器为例来看看电荷分配毛病如何会导致不希望有的结果。图1.55中的 nMOS 移位寄存器由一串反相器组成，反相器之间由钟控传输晶体管隔开，至于使用时钟 ϕ_1 还

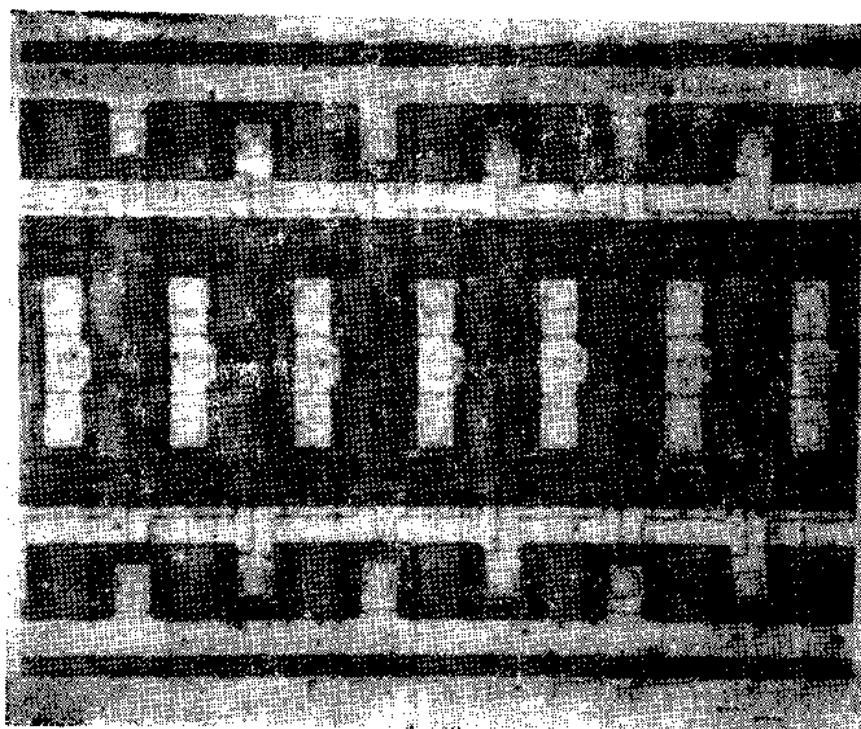


图1.54 采用图 1.52方案的 CMOS 移位寄存器显微照片

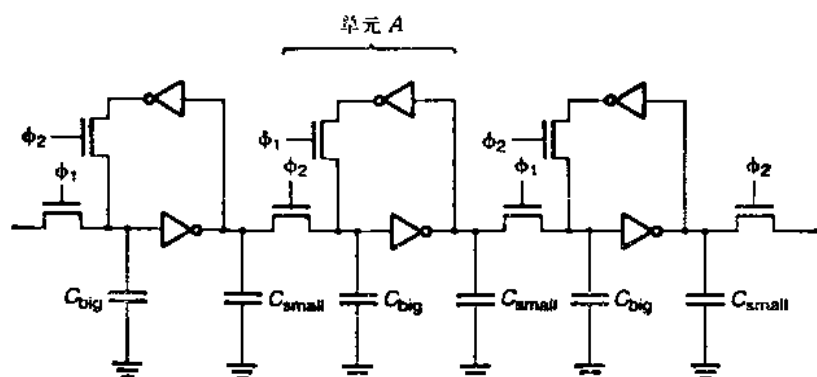


图1.55 一个有电荷分配毛病的移位寄存器，它有可能反向移位

是 ϕ_2 控制则由它们在反相器串中的位置决定。为了使移位寄存器在两个时钟都不为高电平的时候能停顿任意长的时间，设计中加了钟控的反馈回路。由于不幸的巧合，设计中的布线寄生电容使得移位寄存器单元的输入电容比输出电容大得多。这些电容在电路图中分别标为 C_{big} 和 C_{small} 。我们观察一下单元A，当 ϕ_1 变高时，这一单元的反馈回路接通，使得它变为静态的。由于通向下一单元的传输晶体管也通了，所以下一单元的输入也受单元A输出的驱动。不过，若传输晶体管很宽，则当 ϕ_1 变高时发生的第一件事是电荷在 C_{big} 和 C_{small} 之间的分配。由于大小的不同，电容 C_{small} 上的电压将由 C_{big} 上的电压决定。也就是说，这时单元A的输出被一个低阻抗所驱动，并且有可能使单元翻转。电路设计上的一点细节甚至有可能使图1.55中的移位寄存器反向移位。我们还将看到更

多的电荷分配的例子，既有有害的也有有利的。例如，图1.52所示CMOS移位寄存器有电荷分配问题，而图1.48中不那么紧凑的移位寄存器则无此问题。

例1.4 每 N 个时钟周期输出一个脉冲的同步有限状态机的设计。

该问题有很多种解决办法。如果 N 小，我们可以用一个可复位的移位寄存器在循环移位一圈时输出一个信号，如图1.56所示。若 N 大，用这种办法效率就低了，因为它要求移位寄存器的长度为 N 。

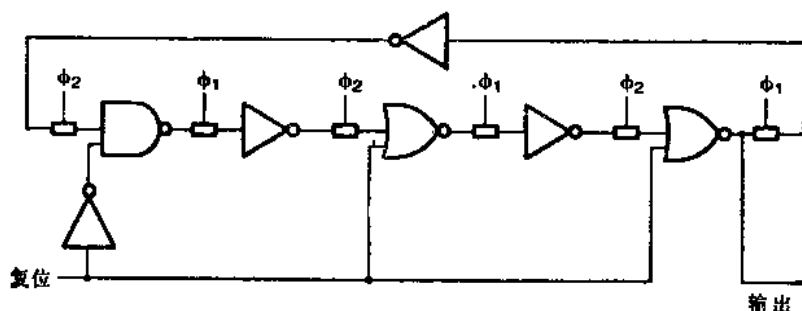


图1.56 一种移位寄存器实现的分频器

第二种解决办法是使用二进制计数器， N 分频只需 $\log_2 N$ 个单元。一种用MOS技术实现的计数单元如图1.57所示。若 T 为“1”，则单元翻转，否则它将不变。每个周期，动态节点被刷新一次。图1.58所示电路说明翻转触发器如何用于计数。我们注意，若所有

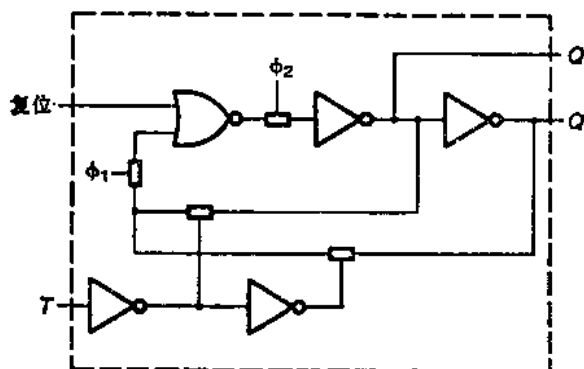


图1.57 一个翻转触发器

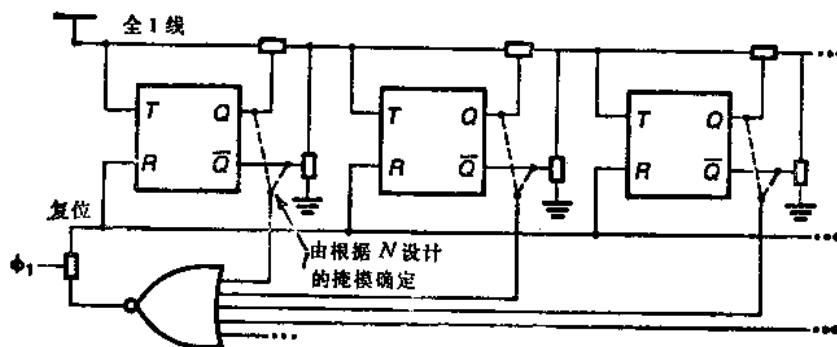


图1.58 二进制计数器

低位均为 1，该单元将翻转，并且只有在这种条件下，该单元才会翻转。“全 1”线就用于传输低位 (LSB) 的这种信息。一个或非门在 N 个时钟信号之后检测正确的数据图形，以便将计数器复位。在实际运用中，或非门往往分布于整个结构，如图 1.59 所示。

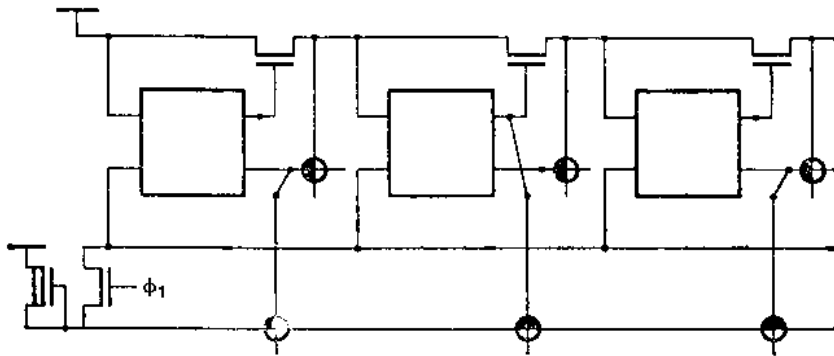


图 1.59 利用分布或非门的二进制计数器

二进制计数器的优点是计数结果很容易辨认。但其关键通道太长。在最坏情形下，每个单元都要一个接一个地翻转。另一种类型的计数器是最大长度计数器或称伪随机顺序计数器，如图 1.60 所示。一个 M 位最大长度计数器将容纳长度为 $2^M - 1$ 位的数据图形 [Dixon 76]。计数的结果不能直读，但对该种应用这并没有多大关系。我们只需要知道数据图形是怎样的就行了。像在二进制计数器的情形那样，我们需要一个给计数器加载的方法。现在计数器仅仅含有一个移位寄存器和一个异或门。由于最大长度顺序计数器的关键通道比二进制计数器短得多，因此它有潜力达到高得多的速度。而且它也比较紧凑。

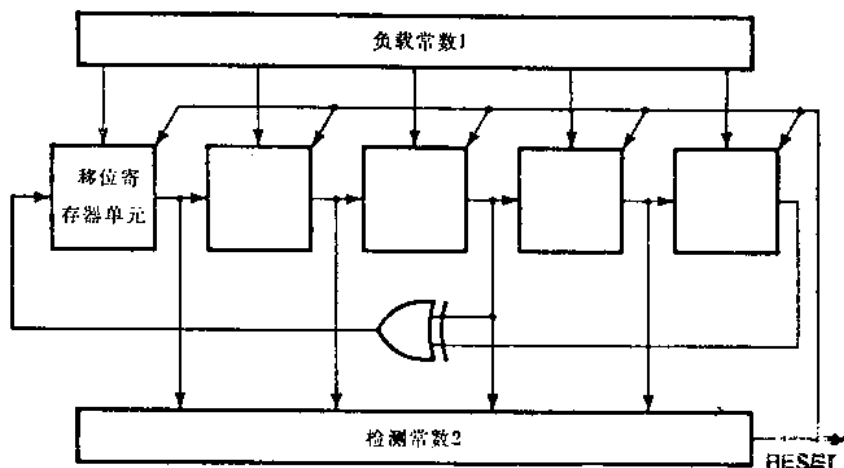


图 1.60 最大长度顺序计数器

预充电

无比 nMOS 电路最普通的应用是预充电。在预充逻辑门或母线中，仍然有上拉及下拉网络。不过，它们分别在不同时间起作用。图 1.61 所示为一条 MOS 母线，它在 ϕ_1 时预充电，而在 ϕ_2 时有条件地放电。这种安排有两个好处。首先，由于从 V_{DD} 到地没有直流

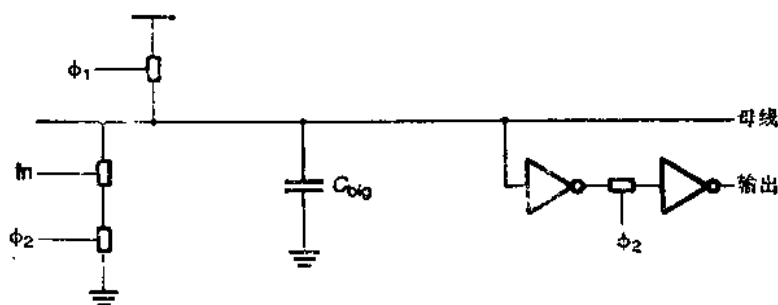


图1.61 一条预充母线

通道，因此这种网络的功耗很低。其次，这种电路的速度大大提高了，这是因为晶体管可以做得比较宽而不会增加功耗，而且输入信号不会使母线上拉。换句话说，在有比的nMOS门中，向上过渡慢而向下过渡相对较快。在预充电路中，若输出应为高，则它在此前已经是高的了，因而这时不需要做什么。这一动作是十分快的。当输出应为低时，母线被一个大的晶体管下拉。这一动作也是很快的。这些都是基于一个假说：系统的结构允许母线有一段时间（本情形中是 ϕ_1 期间）不被使用。而且输入不能有假信号。注意，在 ϕ_2 期间，输入一旦变高一下，母线就将被放电。

在CMOS设计中，预充电路也是极为有用的，这不是为了功耗而是为了速度及版图的紧凑。考虑扇入为 q 的或非门的情形。没有预充电，或非门需要 $2q$ 个晶体管，其中 q 个P型器件要串联在一起。不但这些器件要占很大面积，而且速度也慢，这是因为负载电容要通过高阻抗路径来充电。预充方案则只需 $q+1$ 个晶体管，其速度最多是受一个n沟晶体管限制。这两种电路示于图1.62，其 $q=5$ 。

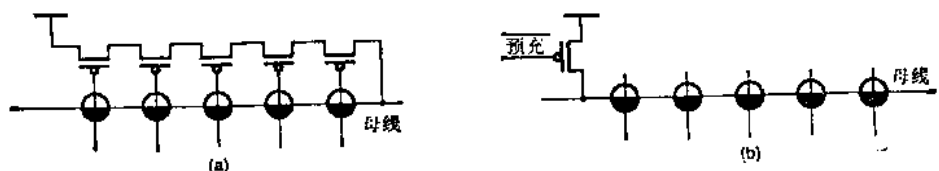


图1.62 (a) 静态 CMOS 母线，(b) 预充电型 CMOS 母线

1.4 规则结构

对于具有十万个晶体管以上的VLSI，如果其最大的单元块只是逻辑门，则它的设计会变得十分冗长。幸而，已经发明了许多种大型规则逻辑结构，它们可以做成集成电路形式。这些结构的一个好处是它们可以增加设计的规则性因子[Lattin 79]。规则性因子是芯片上实际的晶体管数与个别设计的晶体管数之比。显然，大的规则性因子有利于提高设计效率，它对版图也有利。经验表明，在随机逻辑、规则结构与只读存储器中，每个晶体管所占面积之比为10:4:1。不幸的是，哪种规则结构（如果有的话）最适于用来完成一个具体的逻辑功能，并不总能简单地确定。例如，尽管只读存储器（ROM）是集成电路中最紧凑的结构，但如果用ROM来计算一个32位数的奇偶性，两边的尺寸将超过1米。有计算奇偶性更好的办法，其中有些也是规则结构。电路设计者的艺术之一就

是正确选择适合于任务的结构。典型的规则结构有只读存储器，随机存取存储器 (RAM)，可编程逻辑阵列 (PLA)，译码器及移位寄存器等。其它还有许多种。图1.63至图1.66画出了几种形式的MOS译码器电路。第一种方案是基于或非门，第二种与第四种是基于与非门。第三种是一种复杂门方案。

一个ROM可以由一个译码器、一个ROM阵列和一个选择器构成，如图1.67所示。ROM阵列的输出如插页般送入选择器以尽量减少引线。在VLSI中，引线的代价即使不比晶体管大，至少也与晶体管的一样大。

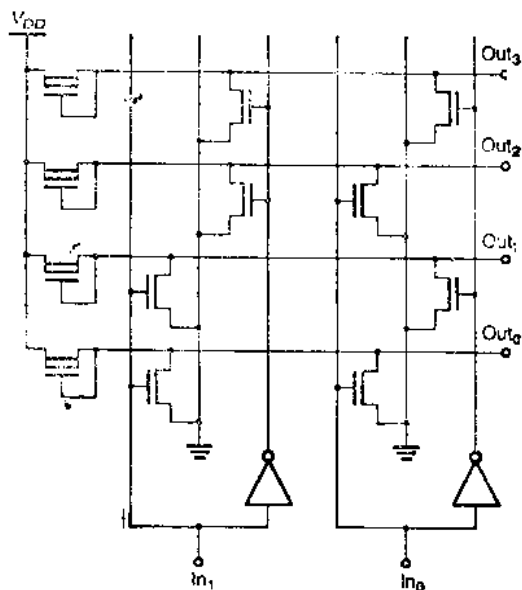


图1.63 或非门形式的译码器

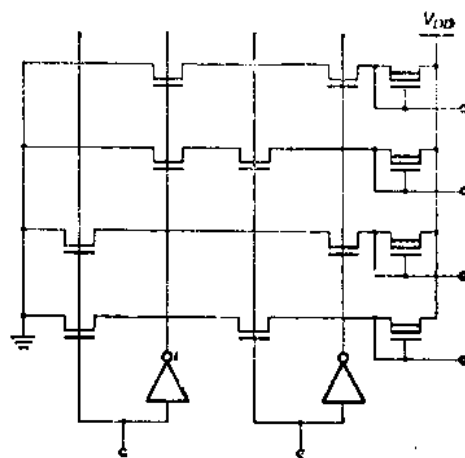


图1.64 与非门形式的译码器

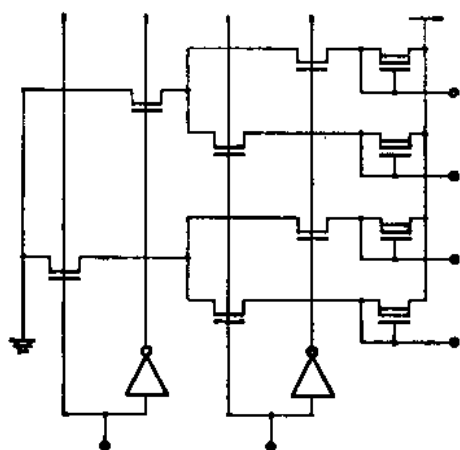


图1.65 伞形译码器

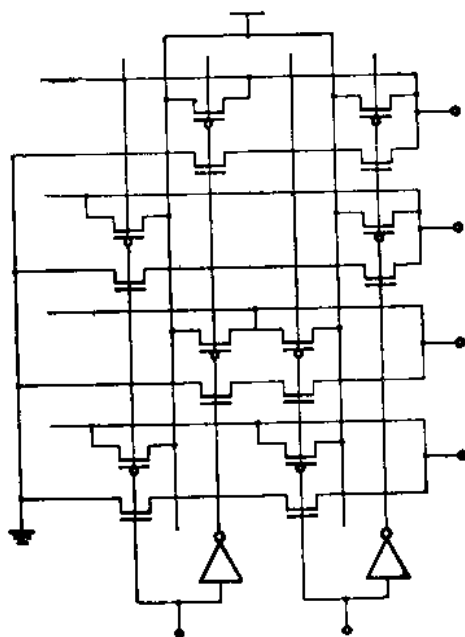


图1.66 CMOS与非门形式的译码器

图1.67所示的ROM有 R 个输入， P 个输出。 R 个输入分别进入译码器和选择器。计算连线的数目可以粗略估计出阵列的面积。如果 R 个输入中有 N 个进入译码器，则阵列的面积 A 粗略地正比于

$$A \propto N2^{N+1} + P2^{N+M} + PM2^{M+1} \quad (1.14)$$

其中

$$N + M = R \quad (1.15)$$

面积极小值发生在

$$1 + N \ln 2 = P [1 + (R - N) \ln 2] 2^{R-2N} \quad (1.16)$$

处，在 $P = 1$ 的特殊情形，可得 $N = M = R/2$ ，最佳阵列是方形的。在其它大多数情形，最佳阵列也几乎是方形的，因为最佳形状对 N 极敏感。若 N 增加1，则阵列的高度要增为2倍，而宽度减为1/2。另一种考虑是 N 与 M 的分配要看如何适应芯片的安排。 M 个选择器信号发生作用的时间要比 N 个译码器信号晚得多。在第八章

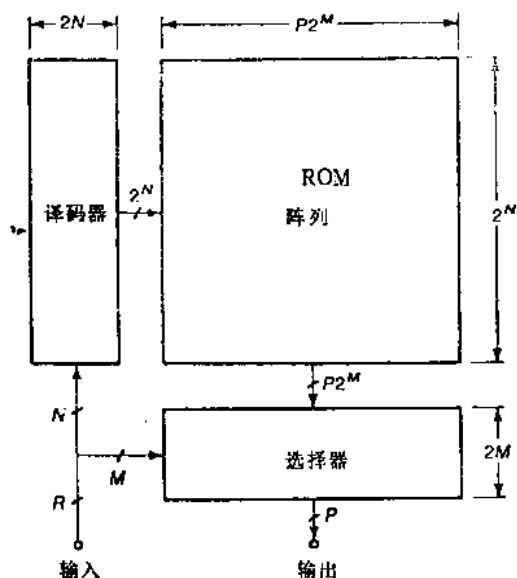


图1.67 ROM阵列

中，我们可以看到在设计一个控制存储器时如何利用这一事实取得好处。

图1.68所示为ROM的一个简单nMOS方案。注意，ROM的“个性”仅仅体现在ROM阵列中晶体管的位置不同。这一点很有用，使得我们可以把ROM的编程放在设计的后期。它也使得工程变动比较容易。

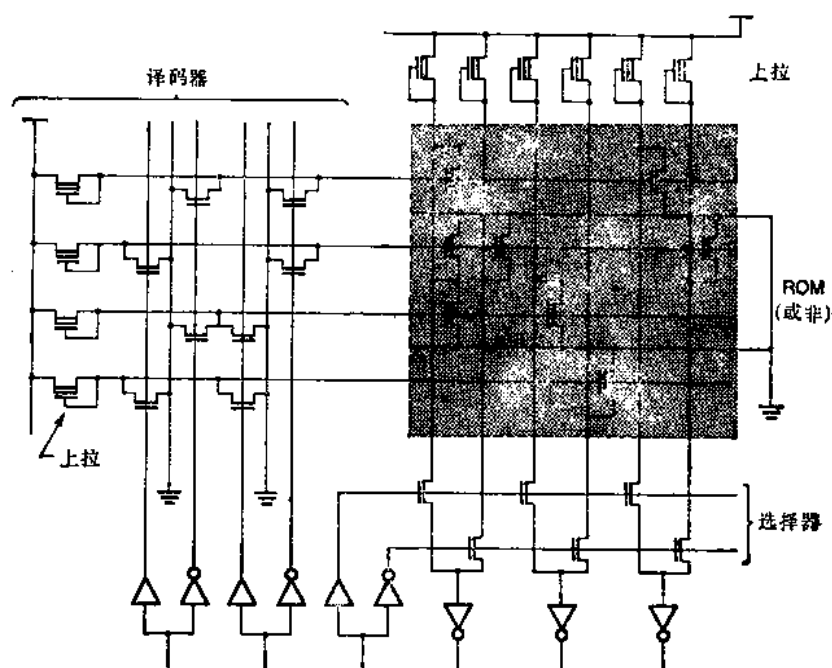


图1.68 一个nMOS ROM方案

图 1.69 一个用 nMOS PLA 实现的全加器

[illegible]

图 1.70 具有串联与阵列的单相PLA。其速度慢但较紧凑

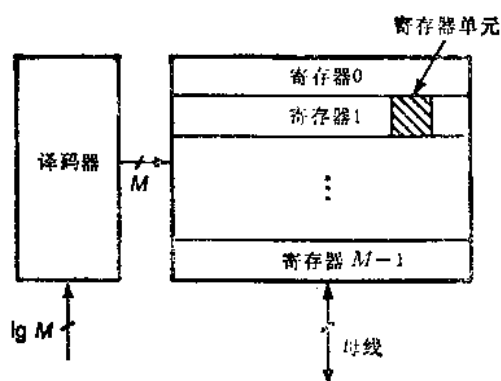


图1.71 一个 M 个字的寄存器阵列

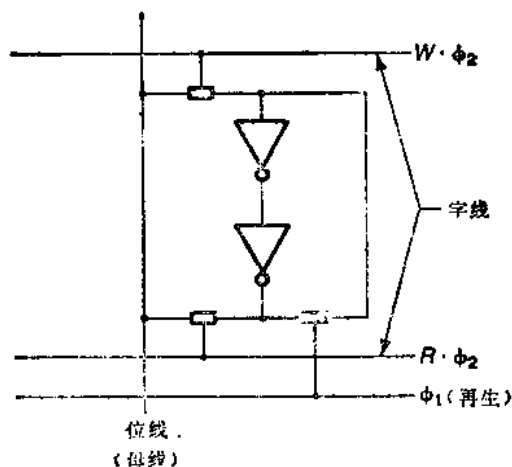


图1.72 一个寄存器单元

变低时，与阵列进行定值。与阵列的输出辗转传入或阵列使或阵列定值。注意，并非与阵列所有节点都需要预充。不过在这种电路中不存在电荷分配问题。实际上，几乎所有用于 VLSI 的大型规则阵列都是预充的。

存储阵列在 VLSI 设计中到处可见。一个存储阵列至少应包括一个地址译码器和一个寄存器单元阵列。图1.71示出了人们在典型微处理器中常见的那种寄存器堆。寄存器单元有许多种可能的构成方法，图1.72为其中之一。这种寄存器堆有效地利用了预充的母线以加速存储器一个字的读出。一次读出是通过在 ϕ_1 期间把母线预充到高电平，而在 ϕ_2 期间把适当的字连接到母线完成的。如果单元存的是高电平，母线将被放电，否则，它将保持预充状态。写入操作也发生在 ϕ_2 期间。在 ϕ_1 期间，单元中的两个反相器交叉耦合从而使单元再生。这样，在 ϕ_1 期间，单元是静态的，而当 ϕ_1 为低时，单元是动态的。因此，仅仅当 ϕ_1 保持高时，系统才能无限期地停顿。这种情况称之为半静态。通过增多晶体管，增加一条或更多条母线，单元可以做成多端的，以备读、写或二者兼之。图1.73所示为一个多端寄存器单元。

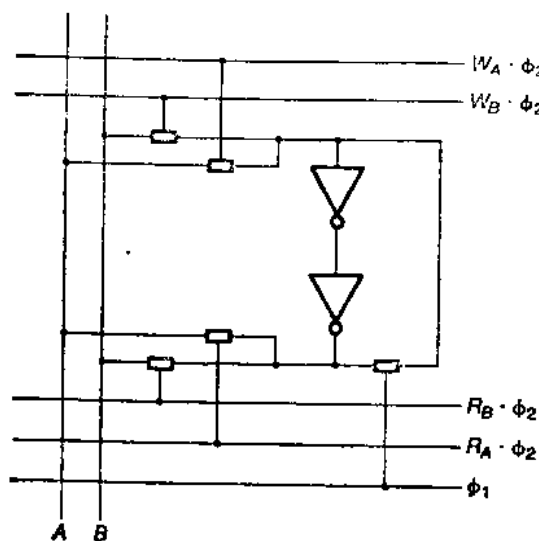


图1.73 一个多端寄存器单元

通过增多晶体管，增加一条或更多条母线，单元可以做成多端的，以备读、写或二者兼之。图1.73所示为一个多端寄存器单元。

逻辑电路与存储器可以交错混合以建立大型的规则结构。这种混合尤易发生在 VLSI 范畴。例子包括存储器、流水线乘法器、卷积器、收缩阵列 [Kung 79]，可按内容寻址的存储器，及多路处理器。在这一领域，大有创新的机会。

例1.5 一个简单乘法器阵列。

阵列乘法器 [Dadda 65] 特别适合于用 MOS 技术来实现。给定两个 M 位数 A 和 B ，我们可以把乘积 $P = A \times B$ 表示为

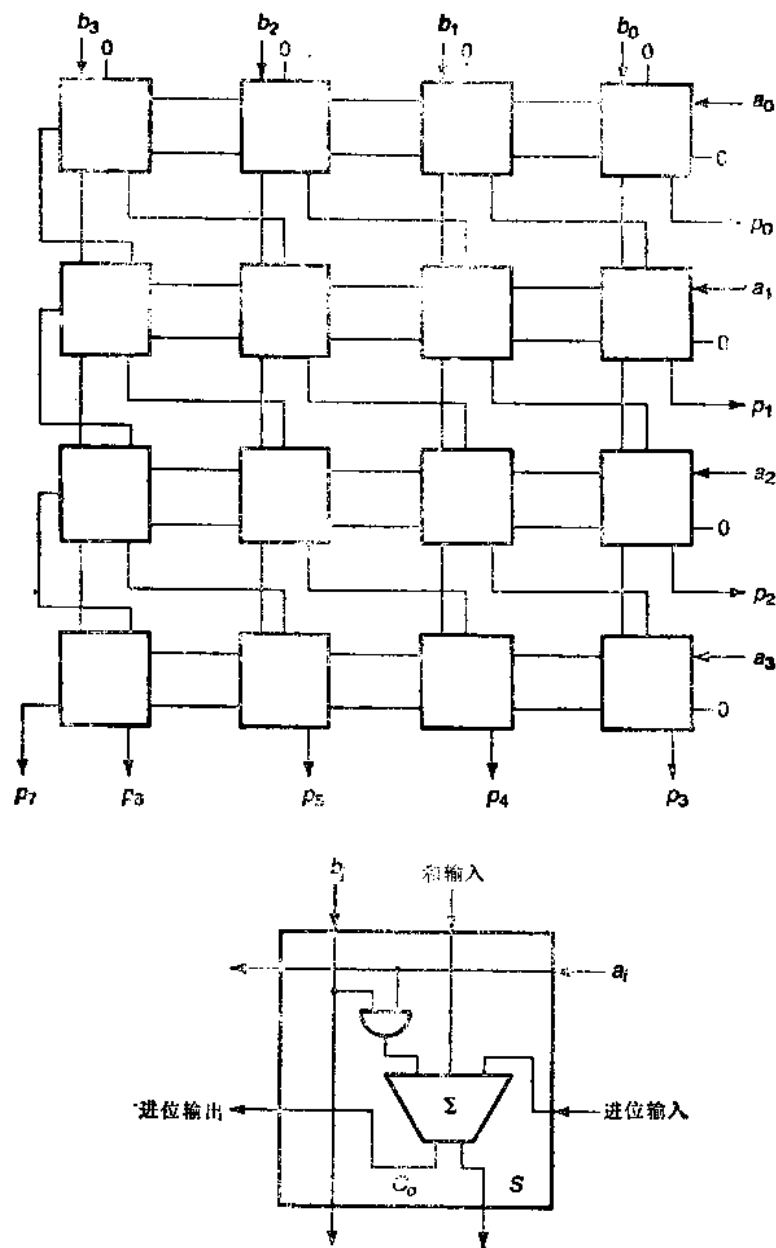


图1.74 阵列乘法器

一串部分积 X_i , 我们有

$$X_{i+1} = X_i + a_i B 2^i$$

其中 $P = X_m$ 。在阵列的第 $i + 1$ 行, 若 $a_i = 1$, 则在部分积 X_i 上加以移 i 位以后的 B 。否则 i 位部分积结果直接转移过来。MOS 乘法器的框图示于图 1.74。这一乘法器在其每一行都采用进位传播加法器。这意味着对每次乘法, 进位可能需要传播 $2M$ 位。在任何一行都可能发生 M 个进位。图 1.75 所示为这种乘法器的一种流水线方案。不幸的是在任何一行最大进位数为 M , 因此时钟周期需要很长。换一种办法, 我们可以在每一

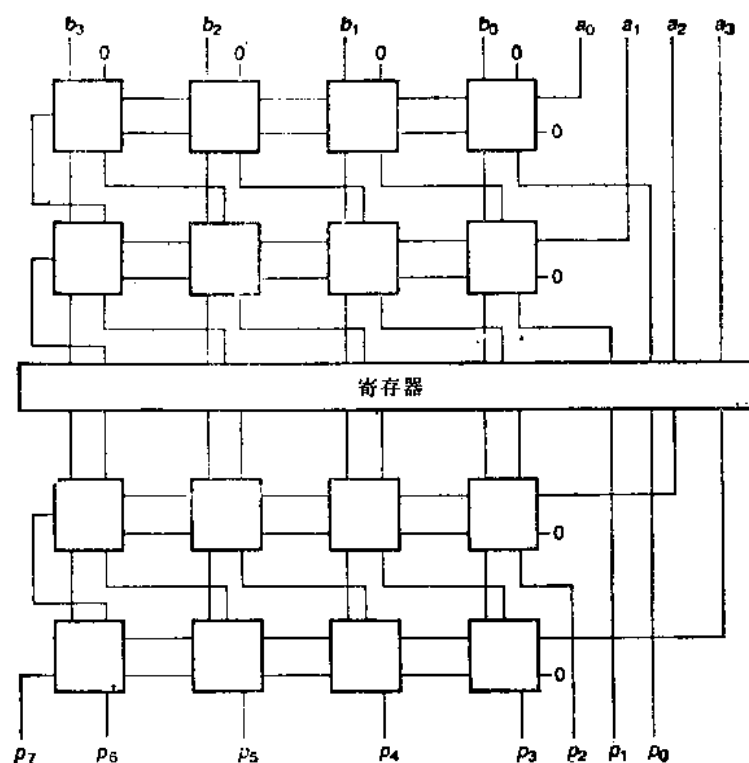


图1.75 用进位传播加法器阵列实现的流水线阵列乘法器

行加更多的逻辑以加速进位，例如采用“先行进位”的技术，但这样做要增加许多面积，因为每一行都要受影响。

一种比较好的技术是除了最后一行外，在所有行采用保留进位加法器。一个保留进位加法器把三个数的和减少为两个数的和，与此同时它将一个信号通过一个全加器单元传播下去。将此与进位传播加法器对比，后者把两个数的和减少为一个数，但其延迟时间正比于被加数的位数，除非采用先行进位的技术或其它一些硬件增强的方案。图1.76给出用保留进位加法器实现的流水线乘法器阵列。同样的基本单元被再次使用。现在，最大进位传播仅是最后一级的问题，这里可以用一个先行进位加法器或类似的方案来打开这一瓶颈。注意仅仅阵列的最底下一行需要先行电路，因此对版图的影响是轻微的。

还可以进行若干附加的优化。例如，使前面所示的阵列的所有单元都是一样的。这样设计与布局布线都比较容易，但如果我们愿意按具体需要作点小调整，就可以节省相当一块面积。例如在保留进位方案中阵列第一行的单元就可以简化为与门。在实际的方案中，将用修正的布局重新编码(Rubenfield 75)使行数减半，不过要增加一些附加的硬件。

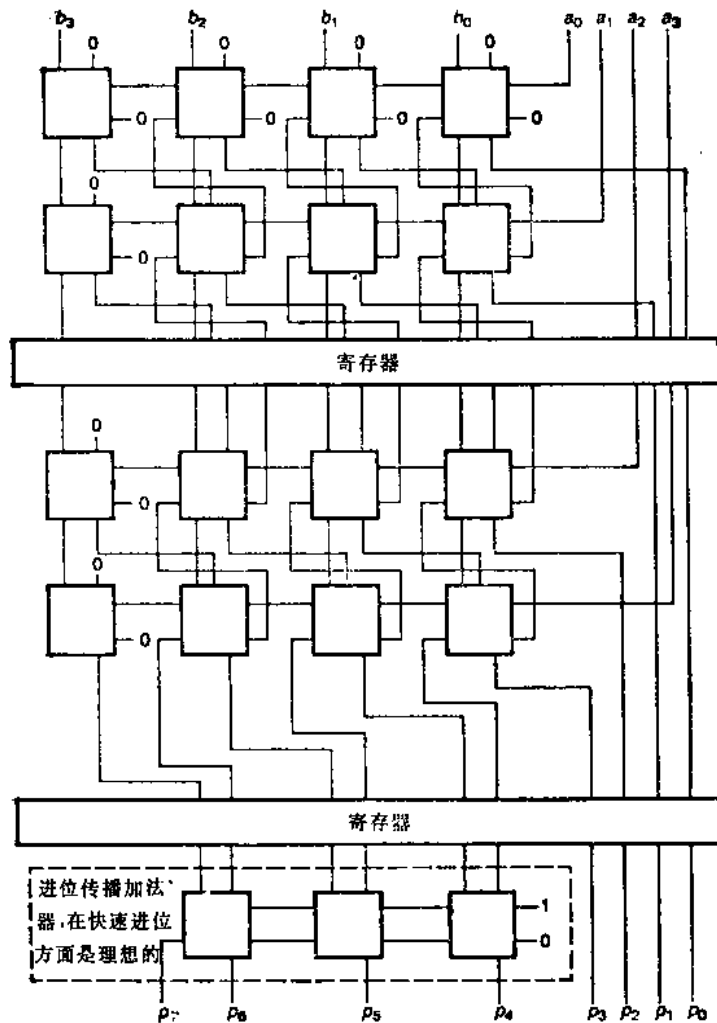


图 1.76 用保留进位加法器列实现的流水线乘法器阵列

1.5 微系统设计

逻辑门与寄存器是 VLSI 系统工程的基本单元。简要地说，VLSI 的功用是变换数据。图 1.77 画出了一个 VLSI 机器，它从一端取入数据，加以操作，从另一端送出。我们所画的称之为数据通道。

数据通道的结构取决于数据的类型、所需的操作，以及性能指标。我们在例 1.5 中看到的乘法器阵列是简单数据通道的一个例子。编为二进制字的数据一般可用以代表数、阵列、文字串、图形，或者无论什么样的对象。

当要对数据进行算术和逻辑操作时，常常用算术逻辑单元（ALU）。典型的 ALU 有两个输入端和一个输出端。对于算术运算，上述任何一种加法器方案都可以被采用。典型的逻辑操作包括每位的与、或、异或等等。要做一个非常普遍的机器只需很少的几



图1.77 一个数据通道

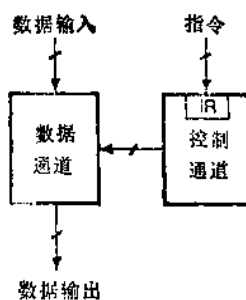


图1.78 一个简单的机器

种基本操作。为了性能方面的考虑，在数据通道要加一些特殊的功能。典型的特殊功能包括移位及查表。移位是一种可以由传输晶体管有效实现的功能，而查表可用 PLA 或 ROM 实现。对于复杂的操作，常常需要变量的暂存器。这方面通常采用寄存器。当预计需要几个寄存器时，寄存器往往摆成一个称之为寄存器堆的阵列形式。

除非数据通道的操作极为简单与重复（像乘法器的例子中那样），都需要一个明确的控制机器来控制数据通道。如果控制序列不是太复杂，一个简单的有限态机¹⁾可以用于实现控制通道。在实现控制通道方面，PLA 是很常用的。大型的 PLA 和 ROM 在控制通道中被称为控制存储器。我们往往需要一定程度的可编程性。一种可能的办法是向控制机送指令。来到控制机的指令往往被存入一个暂存器中，该暂存器被称之为指令寄存器 (IR)，如图1.78所示。

要完成基本的程序机，还需加上存储器，如图1.79所示。控制机的输出现在不仅包括数据通道的控制信号，而且包括两个存储器的存储地址。控制单元中有个专门的寄存器，称之为程序计数器 (PC)，它保持着下一步要执行的指令的地址。给定一个存储器地址，如果给存储器的控制输入为读，则存于该地址的数据被读出，如果控制输入为写，则将数据写入该地址。在上一节对 RAM 的讨论中，我们已经看到了一些简单存储器方案。

要构成一个冯·诺依曼计算机，需要有一种由数据操作结果影响控制通道排序的方式。存储器是共享的。一个经典的冯·诺依曼计算机示于图1.80。

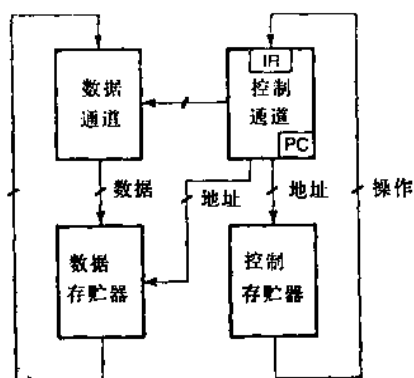


图1.79 一个可编程机器

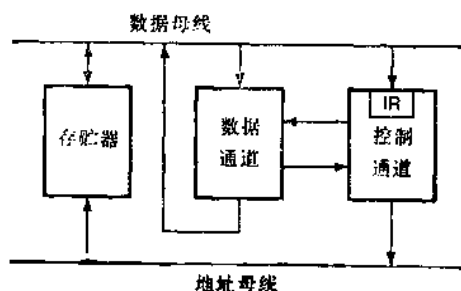


图1.80 一个计算机

1) 有限态机是一种数字机器，它有一个寄存器和组合逻辑反馈通道。在每个时钟周期对输入及反馈项求值，从而产生一个新的输出。

通常, 由于大型存贮器的速度比 VLSI 逻辑的速度慢, 因而使多路数据线与地址线共用一条母线是经济的。这再现了 VLSI 中连线是很昂贵的这一论题。注意我们有两种类型的“数据”。一种是进出数据通道的数据, 而另一种数据实际上是指令。对程序进行操作就像它们是数据一样, 这种能力是冯·诺依曼的根本力量之一。开拓一下我们的思路并且把数据也看作是指令, 有时是方便的。在这种情形下, 那些实际上是简单数据的指令被称为字面上的指令(数据只在字面上看作是而不是真的理解为指令)。

现在我们有了一种简单的机器, 它从存贮器取出指令和操作数。指令在控制通道得到解释, 由它指挥数据通道对操作数进行操作。通常指令是顺序从存贮器取出的, 但也有时要求用数据通道的计算结果来影响控制流。在控制流程的某点, 我们要进行一次条件转移, 最原始的条件转移可能采取下面的形式:

若 ALU 输出为 0, 不要取存在 PC 中的下一条指令, 而是将 PC 的存数

加 77, 然后在存贮器的该地址取出指令。

注意, 现在控制通道需要有做算术操作的能力。有时是用第二个 ALU, 但往往是用数据通道中的 ALU 作地址计算。当选择后者时, 数据和控制通道就紧密地交错在一起了。

程序分支是相关转移的一个例子, 另一个例子是调用子程序。子程序调用的特点在于必须记住相关位置, 以便从子程序回到原来的程序并恢复正常的控制流程。这要求有附加的控制通道寄存器。在一个一般的寄存器机器中, 寄存器堆中的寄存器既可用于控制通道也可用于数据通道。

当计算机所要求的运算非常复杂时, 控制通道所用有限态机就变得很难设计了。摆脱这一困境的一个办法是用一组有限的几种简单操作来构成这些复杂的操作。这样, 计算机可能需要几个内部的周期来处理输入的每条指令, 但这种内部的操作至少应该是容易理解与执行的。这里我们所做的只是重新编排程序并把它应用于微程序水平。所以我们可以给宏机器(它包括数据通道、控制通道和存贮系统)编程序, 并以控制通道自身作为一个计算机[Wilkes 53], 由它编微程序。通常微程序是存贮在 ROM 或 PLA(控制存贮器)中的。微程序解释宏程序的复杂要求, 并把它化简为一种 VLSI 芯片容易实现的简单操作顺序。

1.6 小 结

VLSI 设计者要涉及的范围从器件物理伸展到计算机体系结构。电机工程与计算机科学的很大部分是反映在微观方面。VLSI 就是这些领域的一个很好的研究课题。本章中, 我们定性地观察了 VLSI 的若干方面。我们介绍了一些概念和例子(它们将在本书中反复出现)。我们还定义了一些符号。

我们已经看到, 连线与相互连系主导着 VLSI 的布局。我们还看到, 为了设计 VLSI 电路, 我们必须发展一些自己加的规则也就是一种方法, 用有限数目的部件组成一个 VLSI 系统。规则结构是 VLSI 方法中一个重要部分。

我们介绍了一种 $2\mu\text{m}$ 工艺, 它将在习题与例子中被反复使用。这一工艺的电学与物理参数列于附录中。有关加法器、ROM、微处理器的一些例子也将在本书中反复出现。

在第二章我们将讨论 MOS 器件物理。在理想的情况下，我们也许能从比较高的层次，如逻辑设计、微结构、时钟方法等表象提取出有关 MOS 器件的细节。不幸的是，这种提取过程从来没有完全成功过。因此为了设计高质量的 VLSI 系统，就必须了解 MOS 器件物理。在第二章我们要讨论的重要课题还有建模和计算机辅助设计。一个大型 ROM 的建模与模拟是一个主要的例子。本书中，关键的方程被标以星号 (*)。

在第三章将讨论 VLSI 芯片的制造，我们的注意力将集中于了解制造中的偏差如何影响制成品。决定电路版图尺寸的设计规则是其中的关键概念。

在设计 VLSI 系统时必须把噪声与制造偏差考虑在内。只有关心这些实际问题才能做出可靠的数字机器。如何做到这一点是第四章的主题。我们还将定量地了解一些基本电路形式。我们将看到在第二章所讨论的二级现象，例如漏感应势垒下降，以一级重要性出现在基本电路的情况中。

在第五章，我们将在第二到第四章的基础上讨论多种特殊电路技术，例如输出缓冲器及读出放大器等。我们将再一次看到一些问题的相互影响，例如版图、器件物理、噪声耦合、工艺偏差，以及与相互通讯有关的问题等。

在第六章，我们讨论时钟定时方法。在控制信号在 VLSI 系统中的有序运动方面时钟起核心作用。

在第七章，我们将在第五章的电路和第六章的时钟方法的基础上设计大型的规则结构，例如 RAM、ROM、收缩阵列等。这一章是设计大型系统的铺路石。

在第八章我们讨论 VLSI 系统的微结构。不论好坏，体系结构一直是被技术推动的。我们将讨论宏结构与电路问题、通讯问题相互协调的要求是如何影响 VLSI 系统的性能的。还将讨论在微处理器情况下，与数据通道和控制通道有关的一些关键问题。在一个设计良好的 VLSI 中，各种层次的考虑都应是协调一致的。

习 题

- 1.1 设计一个带有复位的可逆计数器。画出其 CMOS 的电路图。
- 1.2 用一般 MOS 符号画出一个传输晶体管网络的电路图。这一网络可把一个有单个置 1 位的数（如 00010000）转变为它的温度计码（thermometer code）（即 00011111）。
- 1.3 众所周知，任何逻辑功能都可以用或非门单独构成。证明可以只用传输晶体管及反相器得到任何逻辑功能。
- 1.4 用 nMOS 技术，设计一个计算 N 位数奇偶性的网络，所用晶体管不超过 $6N$ 个。
- 1.5 图 1.81 到 1.84 所示的门电路中，C，G，Z 和 V 的逻辑功能是什么。
- 1.6 一个 ROM 需要 2^{19} 位，每次只取 1 位。芯片上留给它的地方是方形的，每边不长于 775 位。首先回答，问题在什么地方，其次，如何在译码器上加一行来解决这一问题？
- 1.7 从图 1.85 到 1.92 的所有移位寄存器都曾在不同场合下应用于实际芯片。给出它们工作所需的假设，描述每一种是如何工作的，画出示意波形，并讨论它们相对的优点和缺点。时钟波形 ϕ_1 与 ϕ_2 是不重叠的。
- 1.8 对于图 1.93 中的电路，当 $V_{in} = 0$ 和 $V_{in} = 5\text{V}$ 时， V_{out} 近似为多少？
- 1.9 图 1.94 中的 V_{out} 假定从地开始，其最大电压是多少？
- 1.10 将图 1.68 的 ROM 一般化后。用 $2^N + P2^M$ 个上拉晶体管（符号见图 1.67），请设计一个方案仅用 $2^N + 2^M$ 个上拉晶体管。

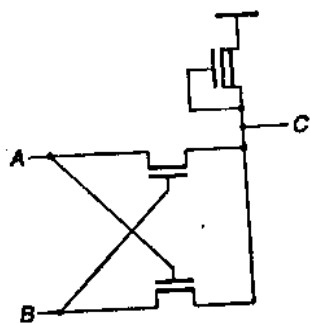


图1.81

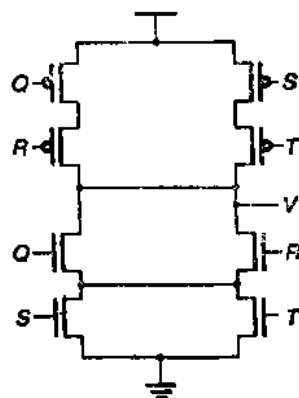


图1.82

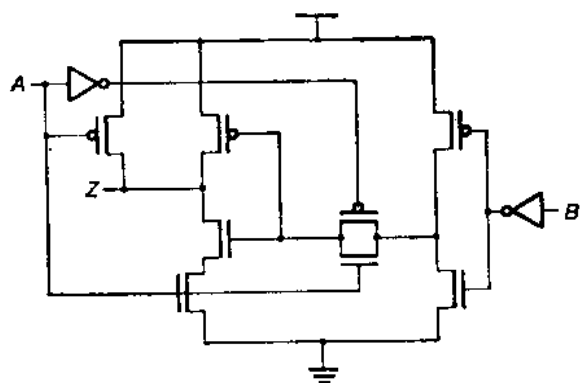


图1.83

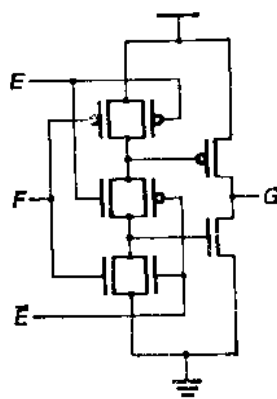


图1.84

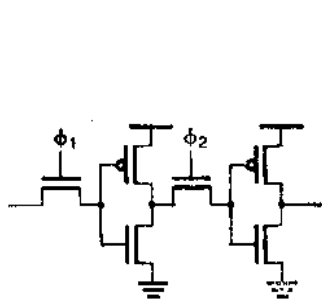


图1.85

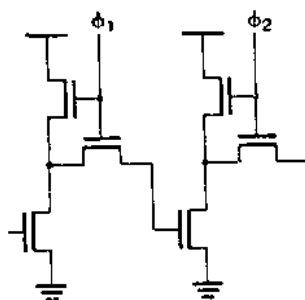


图1.86

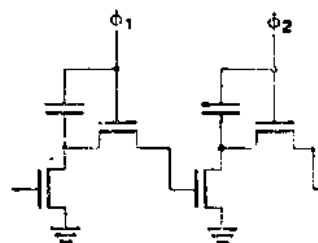


图1.87

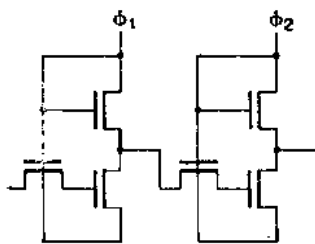


图1.88

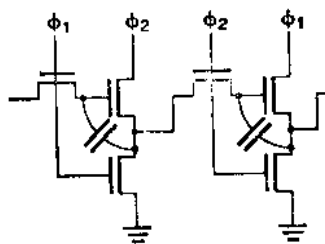


图1.89

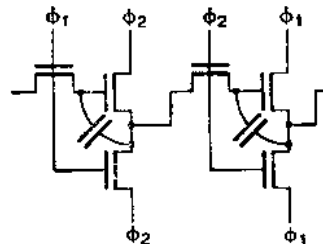


图1.90

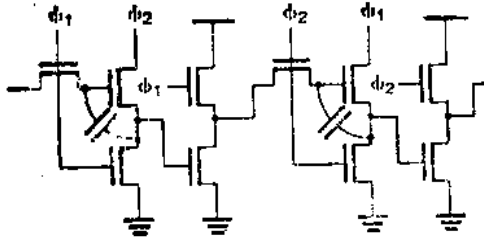


图1.91

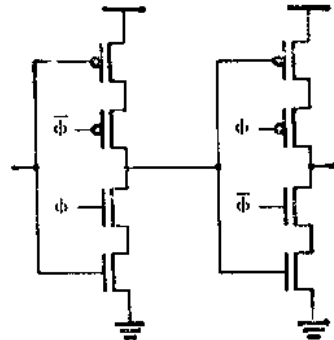


图1.92

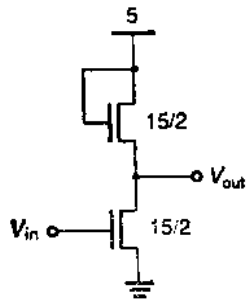


图1.93

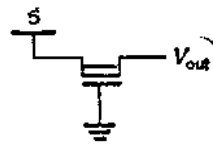


图1.94

- 1.11 设计一个浮点数的语法检验器¹⁾。意思是，给定一个代表字符类型的编码输入后，确定这些字符是否构成有效的浮点数。检验器有三个数据输入 (A_1 , A_0 和启动)，两个时钟信号 (ϕ_1 和 ϕ_2)，以及一个输出“不良” (bad)。数据输入规定了字符的类型，它们按照从左到右的顺序出现。字符类型的信息码如图1.95所示。

A_0	A_1	字符类型
0	0	数字 (从“0”到“9”)
0	1	小数点 (“.”)
1	0	正或负 (“+”或“-”)
1	1	指数符 (“E”或“e”)

图1.95 输入码

图1.96列出了用这一检验器得到的真实和不真实的浮点数。对数字的精度没有限制。时间表列于图1.97。第三个输入 (启动) 在数字的最后一个字符以后和下一个数字的第一个

有效	42	36.	0.7	2E5	2.E0	4.1E+6	-50	5E007	-5.2E2
无效	.4	+.72	+	1.6E	5E+	+.E7	2.4-E3	2..5	5E3.4

图1.96 浮点数字的例子

1) 这个问题是由Steve McCormick提出的。

数 字 流	—	—	3	.	5		+	
输入 (ϕ_1 期间)								
启 动	1	0	0	0	0	1	0	1
A_0	x	1	0	0	0	x	1	x
A_1	x	0	0	1	0	x	0	x
输出 (ϕ_2 期间)								
“不 良”	x	x	x	x	x	0	x	1

图1.97 “- 3, 5”和“+”的输入输出时间表

字符以前至少一个周期才可以是高的。在每个完整的数字后的时钟周期 ϕ_1 ，输出“不良”被取样——一个逻辑“1”表示一个语法上不正确的数。注意在 ϕ_1 期间所有输入都是有效的。

画一个语法检验器的状态图并从这一状态图作一个编码状态过渡表。对下一个状态和输出变量中的每一个，推出最小逻辑方程。

以列表的形式定义实现语法检验器功能的PLA的结构。下面的例子说明了布尔方程 $X = \bar{A}\bar{B} + \bar{A}D$ 和 $Y = \bar{B}CD + \bar{A} + \bar{A}\bar{B}$ 的列表格式。

A	B	C	D	X	Y
1	0			1	1
0			1	1	
	1	0	1		1
0					1

图1.98 PLA列表表示的例子

在图1.98中，“0”或“1”表明晶体管的位置，而空白表示没有晶体管。说明PLA应如何由时钟控制。△

- 1.12 在 nMOS 技术中，实现一个或非门需要 3 个晶体管。在经典的 CMOS 中，由于上拉结构与下拉结构的对偶性，需要 4 个晶体管。不过，如果输入及其补码现成的话，有可能做出 3 个晶体管的 CMOS 或非门，并且不用预充技术和有比逻辑，也不会遭受阈值损失。说明如何做一个 3 晶体管 CMOS 或非门。△
- 1.13 设计一个规则的组合逻辑结构，它可将一个 N 位数解码以找出最低位的零。画出其 nMOS 或 CMOS 的晶体管级的电路图。有 N 个输入和 N 个输出。第 i 个输出线为高表示输入的第 i 位为最低位 0。例如，输入 10010111 将导致输出 00001000。

参 考 文 献

- [Anderson 80] Anderson and Bogert, *LSI Opportunities, The Integrated Circuit Service Industry*, Los Altos, Calif., 1980. (Or its predecessor *Vertical Dis-Integration*)
- [Antognetti 81] P. Antognetti, D. O. Pederson, and H. De man, *Computer Design Aids for VLSI Circuits*, Sijthoff and Noordhoff, Rockville, Md., 1981.
- [Boraiko 82] A. A. Boraiko and C. O. Rear, "The Chip," *National Geographic* 162, 420—457, 1982.
- [Bryant 83] R. Bryant (ed.), *Third Caltech Conference on Very Large Scale Integration*, Computer Science Press, Pasadena, Calif., 1983.
- [Carr 72] W. N. Carr and J. P. Mize, *MOS/LSI Design and Applications*, McGraw-Hill, New York, 1972.
- [Dadda 65] L. Dadda, "Some Schemes for Parallel Multipliers," *Alta Frequenza* 34, 349—356, 1965.

- [Dixon 76] R. Dixon, *Spread Spectrum Techniques*, Wiley, New York, 1976. (Shift register taps of pseudorandom sequence generators.)
- [Gray 81] J. P. Gray (ed.), *VLSI 81*. Academic Press, New York, 1981. (Proceedings of the first International Conference on Very Large Scale Integration, Edinburgh.)
- [Harris n. d.] *HL Cell Library User's Manual*, Harris Semiconductor, Melbourne, Fla., unpublished.
- [Hill 74] F. J. Hill and G. R. Peterson, *Introduction to Switching Theory and Logical Design*, Wiley, New York, 1974.
- [Hodges 83] D. A. Hodges and H. G. Jackson, *Analysis and Design of Digital Integrated Circuits*, McGraw-Hill, New York, 1983.
- [Hon 80] R. W. Hon and C. H. Sequin, *A Guide to LSI Implementation*, SSL-79-7, Xerox, Palo Alto, Calif. 1980.
- [IEEE 78] *IEEE Standard Definitions, Symbols, and Characterization of Metal-Nitride-Oxide Field-Effect Transistors*, IEEE Std. 581--1978, 18, April 28, 1978.
- [Jespers 82] P. G. Jespers, C. H. Sequin, and F. van de Wiele (eds.), *Design Methodologies for VLSI Circuits*, Sijthoff and Noordhoff, Rockville, Md., 1982.
- [Johnston 82] M. Johnston and C. O'Rear, "Silicon Valley," *National Geographic* 162: 458--477, 1982.
- [Joynson 72] R. E. Joynson, J. L. Mundy, J. F. Burgess, and C. Neugebauer, "Eliminating Threshold Losses in MOS Circuits by Bootstrapping Using Varactor Coupling," *IEEE J. Solid-State Circuits* SC-7, 217--224, 1972.
- [Kung 79] H. T. Kung, "Let's Design Algorithms for VLSI Systems," *Proceedings of the Caltech Conference on Very Large Scale Integration*, C. L. Seitz (ed.), Pasadena, Calif., 55--90, January 1979.
- [Kung 81] H. T. Kung, B. Sproull, and G. Steele, *VLSI Systems and Computations*, Computer Science Press, Pittsburgh, Pa., 1981.
- [Lattin 79] B. Lattin, "VLSI Design Methodology: The Problem of the 80's for Microprocessor Design," *Proc. of the Caltech Conference on Very Large Scale Integration*, C. L. Seitz, Pasadena, Calif., 248--252, 1979.
- [McCarthy 82] O. J. McCarthy, *MOS Device and Circuit Design*, Wiley, New York, 1982.
- [Masuhara 72] T. Masuhara, M. Nagata, and N. Hashimoto, "A High-Performance n-Channel MOS LSI Using Depletion-Type Load Elements," *IEEE J. Solid-State Circuits* SC-7, 224--231, 1972.
- [Mavor 83] J. Mavor, M. A. Jack, and P. B. Denyer, *Introduction to MOS LSI Design*, Addison-Wesley, Reading, Mass., 1983.
- [Mead 80] C. A. Mead and L. Conway, *Introduction to VLSI Systems*, Addison-Wesley, Reading, Mass., 1980.
- [Muroga 82] S. Muroga, *VLSI System Design*, Wiley, New York, 1982.
- [Olsen 81] R. E. Olsen and D. W. Dobberpuhl, "A 13,000 Transistor NMOS Microprocessor," *IEEE International Solid-State Circuits Conf.*, 108--109, 1981.
- [Penfield 82] P. Penfield, Jr. (ed.), *Proc. Conf. on Advanced Research in VLSI*, Cambridge, Mass., 1982.
- [Rubenfield 75] L. P. Rubenfield, "A Proof of the Modified Booth's Algorithm for Multiplication," *IEEE Trans. Computers* C-24, 1014--1015, 1975.
- [Rubinfeld 82] P. I. Rubinfeld, "Two-Chip Supermicrocomputer Outperforms PDP-11 Minicomputers," *Electronics*, 131--136, McGraw-Hill, New York, December 15, 1982.
- [Scientific 77] *Scientific American* 237, (3), September 1977.
- [Seitz 79] C. L. Seitz (ed.), *Proc. of the Caltech Conference on Very Large Scale Integration*, Pasadena, Calif., 1979.
- [Seitz 81] C. L. Seitz (ed.), *Proc. of the Second Caltech Conference on Very Large Scale Integration*, Pasadena, Calif., 1981.
- [Shannon 37] C. E. Shannon, *A Symbolic Analysis of Relay and Switching Circuits*, M.S. Thesis, Massachusetts Institute of Technology, Department of Electrical Engineering, 1937.

- [TI 76] *The TTL Data Book for Design Engineers*, (2nd ed.), Texas Instruments, Dallas, Tex. 1976.
- [Vladimirescu 80] A. Vladimirescu and S. Liu, "The Simulation of MOS Integrated Circuits using SPI-CE2," Memo. UCB/ERL M80/7, University of California, Berkeley, October 1980.
- [Vladimirescu 81] A. Vladimirescu, K. Zhang, A. R. Newton, D. O. Pederson, and A. Sangiovanni-Vincentelli, *SPICE Version 2G User's Guide*, University of California, Berkeley, August 10, 1981.
- [Wanlass 63] F. M. Wanlass and C. T. Sah, "Nanowatt logic using field-effect metal-oxide-semiconductor triodes," *IEEE International Solid-State Circuits Conf.*, 32—33, Philadelphia, Pa., 1963. (Early CMOS reference.)
- [Wilkes 53] M. V. Wilkes and J. B. Stringer, "Microprogramming and the Design of Control Circuits in an Electronic Digital Computer," *Proc. of the Cambridge Philosophical Society* 49, 230—238, 1953.

第二章 MOS器件电子学

本章将从电路设计者的角度来考察 MOS 器件的物理问题。我们将说明 MOS 电容及 MOS 器件的各种物理机制是如何在这些器件的外特性中显示出来的。我们的分析将主要针对 n 沟道器件, 这是因为若同时分析 n 沟道和 p 沟道器件极易引起混淆, 但总结时两种器件的结果都将列出。我们还将考察错综复杂的互连布线的物理性质, 这些互连线占据着 VLSI 芯片的大部分面积, 而且往往支配着芯片的性能。

掌握器件物理的基本知识是有助于设计电路的, 而在飞速发展的 MOS VLSI 领域, 为了设计高性能或高可靠性的系统, 这种知识就是绝对必需的了。新出现的技术的特点总是未能很好地概括, 我们要在一种技术过时之前使用这一技术进行设计, 就只有靠对这些新器件基本性质的了解。

这并不是说就没有概括器件电子学的简化方法的余地了, 但是一个利用这一方法的设计者必需懂得潜伏在这种概括下的暗流是什么。

2.1 电容器和二极管

MOS VLSI 中两种最基本的器件是平行平板电容器及反向偏置 $p-n$ 结二极管。与后面我们将讨论的其它器件不同, 这些两端器件的理论已为人熟知并已写入许多教科书中[Grove 67, Muller 77]。我们将从援引固体物理理论的许多结果开始。

固体物理的一些结果

半导体中电子和空穴的浓度可分别写为

$$n = N_c e^{-(E_c - E_F)/kT} \quad (2.1)$$

和

$$p = N_v e^{-(E_F - E_v)/kT} \quad (2.2)$$

其中, N_c 是提供给电子的导带底态密度, N_v 为提供给空穴的价带顶态密度。 N_c 和 N_v 都与绝对温度的 $3/2$ 次方成比例, 方程(2.1)和(2.2)成立的条件是其指数项大约等于或大于 $3^{1)}$ 。方程中 T 代表绝对温度, k 为玻耳兹曼常数。 E_c 和 E_v 分别为导带和价带能级, E_F 则为费米能级。 E_F 与温度是说明电子及空穴能量统计分布的基本参数。对于在平衡状态下的物体, 其各处温度与费米能级应为常数。从方程(2.1)及(2.2)中可以导出一些在直接描述 MOS 器件电学性质时更有用的参数²⁾。平衡状态下, 纯硅的空穴数与电子数相等, 这是因为每激发一个电子到导带就在价带留下一个空穴。

1) 这是因为, 这里玻耳兹曼统计分布被用来近似实际的费米-狄拉克分布。此外, 在高载流子密度下, 态密度函数将被修正

2) 也有许多别的方法, 不必援引能带理论和费米能级来介绍有关材料。某些读者更多地见到的可能是 Gray 和 Searle 的推导[Gray 69]。

在硅中, N_v 与 N_c 的值差不多相同。参看方程 (2.1) 和 (2.2), 对未掺杂的 (本征) 硅, E_f 一定在 E_v 和 E_c 的中间。令方程 (2.1) 的 n 与方程 (2.2) 的 p 相等, 我们得到

$$E_f = \frac{E_c + E_v}{2} + \frac{kT}{2} \ln \frac{N_v}{N_c} \quad (2.3)$$

通过用杂质原子取代硅晶格原子 (掺杂), 可以改变空穴与电子的比例。硅中用以掺杂的典型杂质是产生空穴的硼和产生导带电子的磷或砷。硼为受主原子, 因为它可以从硅晶格接受一个多余电子 (组成共价键)。磷和砷为施主, 因为它们可将其外层的一个多余电子贡献给晶格。注意, 晶体一直保持电中性。例如, 磷外层的第五个电子仅以几十毫电子伏特的能量被束缚于原子核。在室温下它很容易被离化, 成为可以在晶格中自由游弋的载流子。硼的情况说明起来难一些, 它的外层有三个 (而不是五个) 电子, 由于它周围的硅原子有四个外层电子, 因此它很容易接受一个多余电子。当然, 如果硼接受了一个电子, 那么硅原子就缺少了一个电子。这一空穴 (很像水中的泡) 传播的方向与电子运动的方向相反。

若硅中的空穴多于电子, 费米能级将向价带靠近; 若电子多于空穴, 则费米能级向导带靠近。我们定义本征载流子浓度为 n_i , 按定义, 它对空穴和电子是一样的。重写方程 (2.1) 和 (2.2) 为

$$n = n_i e^{(E_f - E_i)/kT} \quad (2.4)$$

和

$$p = n_i e^{-(E_f - E_i)/kT} \quad (2.5)$$

将方程 (2.4) 与 (2.5) 相乘, 我们得到

$$np = n_i^2 \quad (2.6)$$

其中

$$n_i^2 = N_c N_v e^{-E_g/kT} \quad (2.7)$$

$$E_g \equiv E_c - E_v \quad (2.8)$$

E_g 为禁带宽度。平衡状态下, 方程 (2.6) 对所有半导体, 包括掺杂的或未掺杂的, 都是成立的。对于制造 nMOS 器件的典型 p 型衬底, 其 $p = N_A = 2 \times 10^{16} \text{cm}^{-3}$, $n_i = 1.45 \times 10^{10} \text{cm}^{-3}$, $n = 10^5 \text{cm}^{-3}$ 。 E_g 随温度有微弱变化。

我们用 N_A 表示受主杂质原子浓度, N_D 表示施主浓度。室温下, 实际上所有杂质原子都是离化的¹⁾。当 $N_A \gg N_D$, 而且 $N_A \gg n_i$ 时, $p \simeq N_A$; 当 $N_D \gg N_A$, 而且 $N_D \gg n_i$ 时, $n \simeq N_D$ 。

从电路的观点看, 用电位来讨论问题比用能量方便得多。我们定义费米势为

$$\phi_F \equiv \frac{E_f - E_i}{q} \quad (2.9)$$

其中 q 为电子电荷。注意, 现在我们的费米势是以其本征值为参考的。这是我们将碰到的几种参考电位移动中的头一种。在“F”后面再加一个下标则表示此电势是在 p 型材料 ($N_A > N_D$) 还是 n 型材料 ($N_D > N_A$) 中测量的。对 N_D 或 N_A 为主的情况, 分别有

1) 当温度降至液氮或更低时, 或者杂质浓度极高时, 情况就不是这样了。

$$\phi_{Fn} = -\frac{kT}{q} \ln \frac{N_D}{n_i} \quad (2.10)$$

和

$$\phi_{Fp} = -\frac{kT}{q} \ln \frac{N_A}{n_i} \quad (2.11)$$

最后，我们可将方程 (2.1) 和 (2.2) 重写为

$$n = n_i e^{q\phi_F/kT} \quad (2.12)$$

$$p = n_i e^{-q\phi_F/kT} \quad (2.13)$$

图2.1所示为本征硅、p型硅及n型硅的能带图。

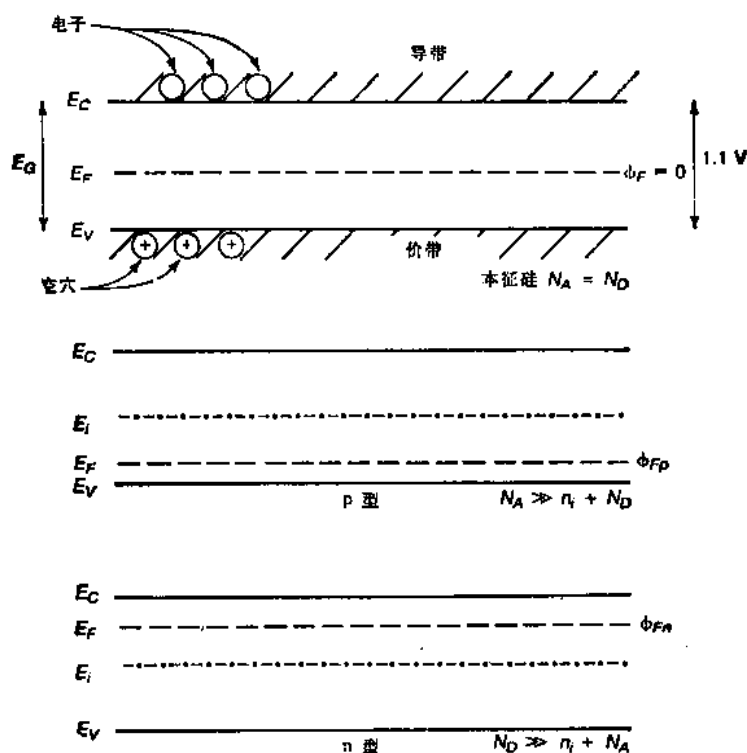


图2.1 硅的能带图。(a) 本征；(b) P型；(c) n型

p-n二极管

p-n二极管在 MOS 电路中到处可见。其电容可能成为影响系统性能的主要因素。

至今为止，我们推导的结果在均匀掺杂的半导体内都是成立的。这些结果在经典的 p-n 结二极管中远离结的区域也是成立的。这样，二极管中 p 区内部的电势就是 ϕ_{Fp} ，而在 n 区一侧电势为 ϕ_{Fn} 。平衡状态下（未加电压），在 n 型和 p 型材料中的费米能级必须相等。从一个区域到另一区域，所观察到的电势总的改变为

$$\phi_T = \phi_{Fn} - \phi_{Fp} \quad (2.14)$$

进一步代入，可得

$$\phi_T = \frac{kT}{q} \ln \frac{N_D N_A}{n_i^2} \quad (2.15)$$

自建势 ϕ_T 来源于 n 型和 p 型材料化学势之差。尽管半导体从整体上看是不带电的，但是 p-n 结的非连续性使得界面附近的材料产生局部体电荷。对于图 2.2 所示一维 p-n 结，泊松方程要求 ϕ_T 等于电荷密度的积分，

$$\phi_T = -\frac{q}{\epsilon_{si}} \iint n(x) dx dx' \quad (2.16)$$

其中 ϵ_{si} 为硅的介电常数， $n(x)$ 为一维电荷密度。

从物理上看，在结处存在着电荷扩散与中和趋势的平衡。界面 p 侧的空穴要扩散到 n 侧，而 n 侧的电子要扩散到 p 侧。这导致电中性的破坏和随之产生的电场。电场的方向与电荷流动的方向相反。在平衡下这两种趋势完全抵消。求得界面附近电荷分布的准确表达式是很困难的。不过，利用耗尽区近似可以得到非常好的结果。所谓耗尽区是指半导体材料中可动载流子耗尽的部分。

在耗尽区近似中我们假定耗尽区在 n 型区一侧延伸的距离为 X_N ，而在 p 型区一侧延伸的距离为 X_P 。根据电中性要求，有

$$qN_D X_N = qN_A X_P \quad (2.17)$$

我们采用了突变 p-n 结模型，其中 n 侧的施主浓度和 p 侧的受主浓度都是常数。电荷密度 $Q(x)$ 为

$$Q = \begin{cases} qN_D & -X_N \leq x < 0 \\ -qN_A & 0 < x \leq X_P \\ 0 & \text{其它地方} \end{cases} \quad (2.18)$$

$x = 0$ 处为 p-n 结的界面。利用

$$\phi_T = -\frac{1}{\epsilon_{si}} \int_{-X_N}^{X_P} \int_{-\infty}^{x'} Q(x) dx dx' \quad (2.19)$$

可求出

$$\phi_T = -\frac{qN_D X_N^2 + qN_A X_P^2}{2\epsilon_{si}} \quad (2.20)$$

与方程 (2.17) 联立，可得

$$X_{PN} = X_N + X_P = \sqrt{\frac{2\epsilon_{si}}{q} \frac{(N_A + N_D)}{N_A N_D} \phi_T} \quad (2.21)$$

这里 X_{PN} 为耗尽区的总宽度。此外

$$\frac{X_P}{X_N} = \frac{N_D}{N_A} \quad (2.22)$$

往往 p-n 结的一侧是重掺杂的，而另一侧是轻掺杂的，这种情况称之为单边突变结。这种情况下，方程 (2.21) 化简为

$$X_{PN} = \sqrt{\frac{2\epsilon_{si}}{q} \frac{\phi_T}{N}} \quad (2.23)$$

其中，当 $N_D \ll N_A$ 时， N 为 N_D ，而当 $N_A \ll N_D$ 时， N 为 N_A 。注意耗尽区宽度主要由

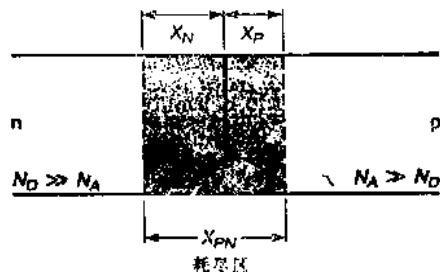


图 2.2 p-n 结二极管

轻掺杂的半导体材料决定。

上述方程是在零偏压条件下推导出来的。在反偏条件下， ϕ_r 项除了方程 (2.15) 中的自建势之外还应包括外加电压。将外加电压 v 包括在内，方程 (2.23) 可改写为¹⁾

$$X_{PN} = \sqrt{\frac{2\epsilon_{si}(\phi_r + v)}{qN}} \quad (2.24)^*$$

从方程 (2.24) 可以看出，大的反向偏置 v 将使耗尽区宽度增加（反向偏置 v 是这样确定的，正的 v 值意味着加到 p 型材料的电位相对于加到 n 型材料的电位为负）。

反向偏置的单边突变 p-n 结的电容可写为

$$\begin{aligned} C &= \frac{dQ}{dv} = \frac{d}{dv} qNX_{PN} \\ &= \sqrt{\frac{\epsilon_{si}qN}{2(\phi_r + V)}} \end{aligned} \quad (2.25)^*$$

按照常规，我们用大写的变量 (V) 代表直流 (dc) 量，而用小写的变量 (v) 代表动态量或大信号。

加到宽耗尽区上的电压，由于雪崩击穿而受到限制。当半导体中的电场足够强时，可动电子在两次碰撞间获得的能量可能大大超过禁带宽度 E_g 。当一个这样的高能电子与硅原子碰撞时，硅原子可能被离化。由于每个电子可以激发出两个附加的载流子（一个电子和一个空穴），电流开始倍增。若载流子产生率超过复合率，就会发生雪崩击穿。单边突变结的击穿电压 V_M 由下式给出：

$$V_M = \frac{\mathcal{E}_M X_{PN}}{2} = \frac{\epsilon_{si} \mathcal{E}_M^2}{2qN} \quad (2.26)$$

其中 $\mathcal{E}_M$ 为雪崩倍增发生时的电场。 $\mathcal{E}_M$ 的经验表达式为

$$\mathcal{E}_M = \frac{4 \times 10^5}{1 - \frac{1}{3} \log_{10}(N/10^{16})} \quad [\text{V/cm}] \quad (2.27)$$

其中 N 的单位为 cm^{-3} [Sze 81, p.102]。

MOS 电容

MOS 电容不仅是 MOS 晶体管的基础，而且它本身也是吸引人的、有用的电路元件。它是一种非线性很强的器件，这种非线性特性在许多电路技术中是十分重要的。

MOS 电容把平行平板电容器和 p-n 结二极管的特性结合在一起。它是由半导体、氧化物和金属组成的多层结构（图 2.3）。在典型的 nMOS 情形，半导体为轻掺杂的 p 型硅，氧化物为热生长的二氧化硅，金属为铝或简并掺杂²⁾的 n 型硅。只要半导体的导电性足够好——能支持位移电流，这种结构就成为一种平行平板电容，在这一例子中，以铝为一个极板以 p 型硅为另一极板，电介质是二氧化硅。定义 C_{ox} 为这种结构单位面积的电容量：

1) 带 * 号的方程表示十分重要。

2) 简并掺杂的意思是，掺杂很浓以致于费米能级到达价带 (p 型) 或导带 (n 型)。这时的半导体几乎和金属一样。

$$C_{ox} = \frac{\epsilon_{SiO_2}}{T_{ox}} \quad (2.28)^*$$

其中 T_{ox} 为氧化物的厚度, ϵ_{SiO_2} 为二氧化硅的介电常数。牢记 C_{ox} 为单位面积的电容是很重要的。

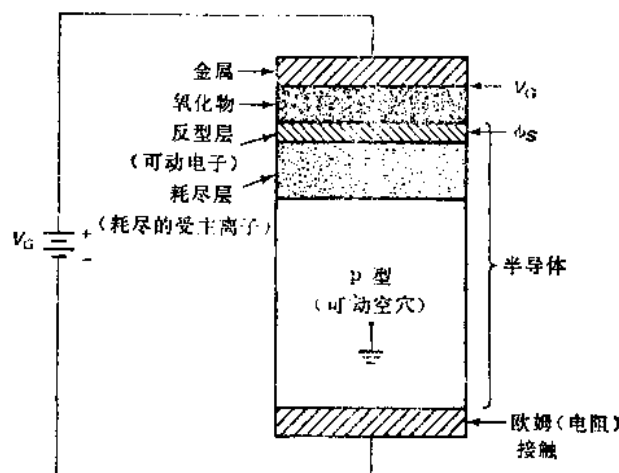


图2.3 MOS电容器

从前面对 p-n 结的分析可知, 硅的化学势与其掺杂程度有关。毫不奇怪, 铝、二氧化硅和 p 型硅也各自有不同的化学势, 而且这些化学势会导致产生自建电场和非均匀电荷分布。其效果是使能带图中的能带弯曲。对于几何形状和杂质分布都非常简单的情况, 我们可以用外加电压来抵消这种能带弯曲, 使半导体呈电中性。达到这一结果所需的电压就是平带电压。这是一个重要的参考电位。图2.4所示为处于平带情况的 MOS 电容器。

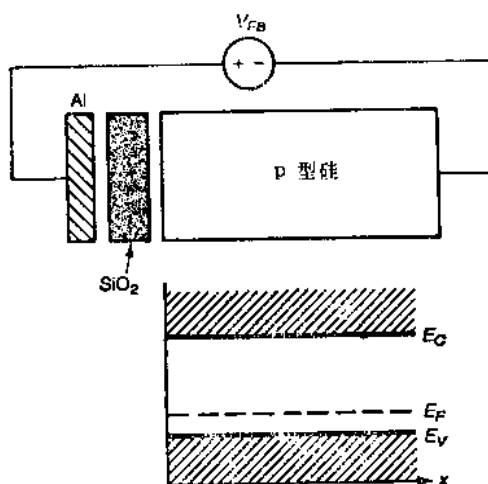


图2.4 平带下MOS电容的能带图

当所加的电压 $v_G < V_{FB}$, 能带的弯曲如图2.5(a)所示。这种情况称为积累, 较低的电压使空穴被从衬底拉出来积累到硅-二氧化硅界面。这种情况下电容不变。

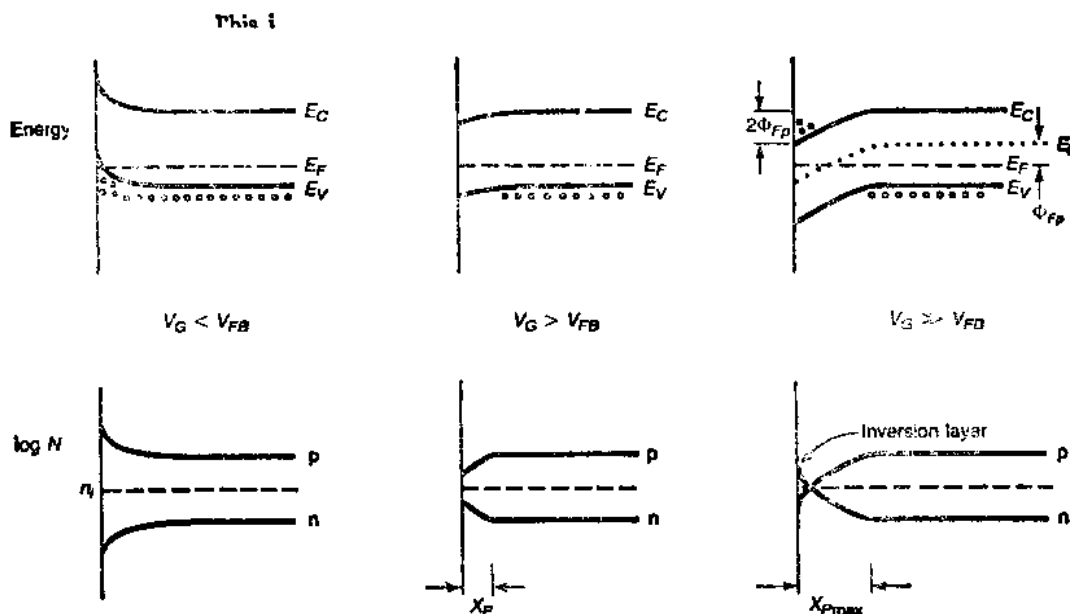


图2.5 P型半导体表面附近的能带图和载流子浓度

当所加电压 v_G 略大于 V_{FB} 时, 器件处于耗尽情况。在这种情况下, 能带向下弯曲如图2.5(b)所示。由于半导体表面的电势高于体内 (ϕ_{fp}), 根据方程(2.13), 其空穴数将减少。这样, 这一区域的空穴开始耗尽。不过 np 值仍保持为常数 n_i^2 。这一耗尽区犹如一个电容, 正像反向偏置的 p-n 结二极管的情况那样。结果是 MOS 电容器的有效极板从氧化层界面下移到半导体体内。这时有两种电介质相串联, 即 ϵ_{SiO_2} 和 ϵ_{Si} 。这使得单位面积的电容比 C_{ox} 小, 它是这两个电容的串联值, 即:

$$C_{MOS} = \left(\frac{1}{C_{ox}} + \frac{1}{C_{si}} \right)^{-1} = \left(\frac{T_{ox}}{\epsilon_{SiO_2}} + \frac{X_p}{\epsilon_{Si}} \right)^{-1} \quad (2.29)$$

其中 X_p 为感应耗尽区的宽度。

若我们增加 v_G 使能带继续下弯, 表面电子的浓度可能超过表面空穴的浓度, 如图2.5(c)所示。当发生这种情况时, 我们说表面进入弱反型。若 v_G 进一步增加, 表面电子浓度将等于并进而超过衬底空穴的浓度。这种情况称之为强反型, 它对电路十分重要。强反型发生时的电压称为阈电压。对强反型¹⁾ 电压的方便的定义是此电压下表面电子浓度等于衬底内部空穴浓度。我们知道, 相对于本征半导体, 衬底内的费米势 ϕ_F 为 ϕ_{fp} 。为了使表面电子浓度与体内空穴浓度相等, 表面势必须为 $-\phi_{fp}$ (由于 ϕ_{fp} 为负值因而 $-\phi_{fp}$ 为正值)。电位差为 $-\phi_{fp}$ 。这是一个非常重要的量。为了使以后对方程的处理更简洁, 我们定义

$$\Phi_{fp} \equiv |\phi_{fp}| \quad (2.30)$$

现在再一次移动参考电位是有好处的。我们要以衬底作为地。这就是说, 我们设定半导体体内电位为零而不是以本征硅的费米势作为参考。这使衬底费米势移动到 $-\phi_{fp}$ 。

1) 此后反型即指强反型。

反型表面电位移移动到正的 $2\Phi_{FP}$ 。这 $2\Phi_{FP}$ 中,一个 Φ_{FP} 用于从p型转为本征,另一个 Φ_{FP} 用于从本征转为n型。反型表面¹⁾的电位为 $2\Phi_{FP}$ 。

反型情况下的MOS电容器的电容比其它两种情况要复杂得多。为了了解这一点,我们必须搞清楚在反型表面处的电子是从哪里来的。一方面,它们来源于电子与空穴的平衡,如方程(2.6)所示, n_p 为常数。关键是“处于平衡”。半导体达到平衡需要多长时间?在硅-二氧化硅界面的电子只能来源于载流子的热产生。在正常工作条件下,这一过程是很慢的(在毫秒量级)。如果我们以很慢的速度测量反型下MOS结构的电容,我们测到的电容值将是 C_{ox} 。这是由于金属极板上的电荷将使由反型层中产生等量的相反电荷。载流子的热产生将使耗尽层电容短路。这种情况如图2.6所示。另一方面,如果我们在高频下进行测量,当 v_G 比发生反型的阈电压大的时候,电容也不回到其最大值 $C_{ox}WL$,但是它也不再继续减少。这是一个微妙的但很重要的特点。当 v_G 慢慢增加,超过发生反型的电压后,所有新增加的 v_G 都降在氧化层上,而硅的表面势固定为 $2\Phi_{FP}$ 。为了说明这一现象,我们可以分析一下电位随界面电荷密度改变的灵敏度。

表面电荷密度改变10%将使氧化层电容两端的电压也改变10%,这是因为 $\Delta v = q\Delta n/C_{ox}$ 。另一方面,表面势 ϕ_s 与电子密度的关系由方程(2.12)决定。在反型时

$$n = n_i e^{q\phi_{FP}/kT} = N_A \quad (2.31)$$

在这一工作点

$$\frac{\partial n}{\partial \phi_s} = \frac{q}{kT} N_A \quad (2.32)$$

n 改变10%, ϕ_s 仅变化2.5mV。这是由于指数关系使得MOS电容器的表面势在 $2\Phi_{FP}$ 以上时,靠增加栅电压已很难使表面势增加,这与正向偏置的p-n结的压降很难超过 ϕ_i 是一样的。硅表面势的微小变化就会使电子数急剧增加。

利用方程(2.23)我们可以写出 X_p 与 ϕ_s 的关系:

$$X_p = \sqrt{\frac{2e_{Si}}{q} \frac{\phi_s}{N_A}} \quad (2.33)$$

其中 X_p 具有极大值:

$$X_{pmax} = \sqrt{\frac{2e_{Si}}{q} \frac{2\Phi_{FP}}{N_A}} \quad (2.34)^*$$

MOS电容器最大与最小电容的比值²⁾为

$$\frac{C_{max}}{C_{min}} = 1 + C_{ox} \sqrt{-\frac{4\Phi_{FP}}{e_{Si}qN_A}} \quad (2.35)$$

其中 C_{max} 为 $C_{ox}WL$ 。通过电容测量可以用这一方程来确定 N_A 。图2.6所示为理想的MOS电容 $C-V$ 曲线。图2.7画出了 C_{min}/C_{max} 与掺杂浓度及氧化层厚度的函数关系[Goetzberger 66]。若 N_A 不是均匀的,电容-电压特性可被用以求得杂质分布。

图2.8所示为电路中MOS电容的一般结构,它有一个侧向的引出端。当这一电容处于反型区时,侧面的n⁺引出端可以向反型层提供电子,而不需要依靠缓慢的热产生过

1) 在2.3节中我们将对这一结果略作修正,这是由于硅表面的掺杂浓度可能与体内不同。

2) 由于 $\phi_s = 2\Phi_{FP}$ 之前器件已进入“弱反型”区,因此实际上该值要稍高一些。

程。这样，直到很高的频率，这种器件的 $C-V$ 特性也与低频的相同。

在阈值电压以下，不存在反型层，因此在栅与侧向接触之间几乎没有电容。栅与体硅之间的电容已示于图2.6。

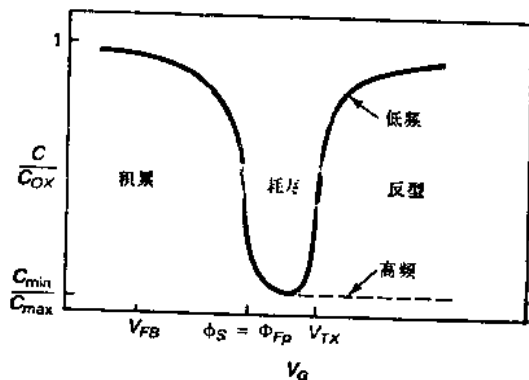


图2.6 MOS电容器的电容特性

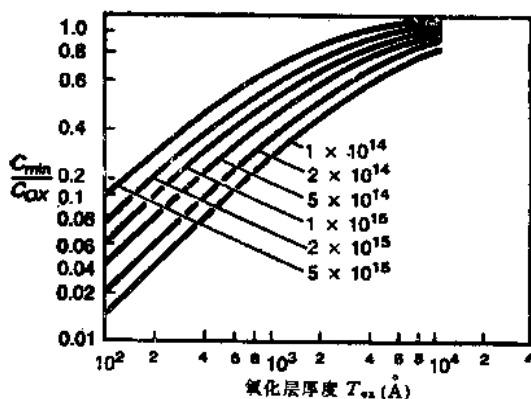


图2.7 不同掺杂浓度下 C_{min}/C_{OX} 与 T_{OX} 的关系[Goetzberger 66 p.1121]

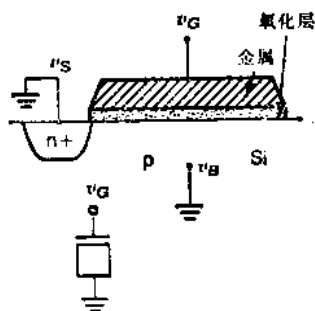


图2.8 电路中的MOS电容结构

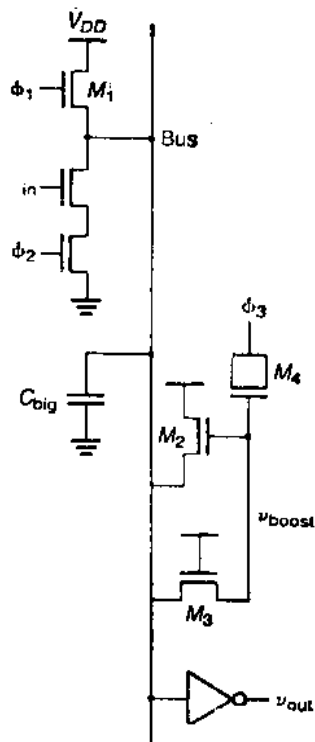


图2.9 利用变容二极管效应的
母线自举电路

例2.1 使MOS电容器单位面积电容降到 C_{min} 以下。

在栅上加一个脉冲大电压可以使电容值降到 C_{min} 以下。在表面载流子产生以恢复平衡之前， ϕ_s 可升到 $2\phi_{fp}$ 以上，这将使耗尽区加厚。这一效应称之为“深耗尽”。

例2.2 用变容二极管泵恢复MOS母线电平。

图2.9所示母线助举电路的作用是，当母线为高电平时将母线电压举高到 V_{DD} 而当

母线为低电平时不起助举作用。当 ϕ_1 来到时, 通过晶体管 M_1 , 母线被预充到 $V_{DD}-V_T$ 。在 ϕ_2 期间输入被采样, 根据输入的状态, 母线或是保持在 $V_{DD}-V_T$, 或是被放电到地电位。在 ϕ_3 期间母线助举电路发生作用。在 ϕ_3 到来前 M_3 是导通的。若母线处于高电平, 则 v_{boost} 被充到 $V_{DD}-V_T$ 。这意味着 M_4 处在阈值以上, 因而栅下存在沟道。这样, 在 v_{boost} 与 ϕ_3 之间就存在大电容, 当 ϕ_3 上升时, v_{boost} 将被泵到 $V_{DD}+V_T$ 以上, 这使得母线可以通过 M_2 被充到 V_{DD} (这里假定 M_4 的尺寸是合适的)。

若母线处于低电平, 则 v_{boost} 被放电到地。这意味着 M_4 处在阈值之下, 因而不存在沟道。这样 M_4 的电容就很小。当 ϕ_3 升高时, v_{boost} 不会升高到 V_T 以上 (这里同样假定 M_4 的尺寸是合适的)。因此母线将停留在低电平。

平带电压

我们已经了解到平带电压在 MOS 器件分析中的重要性。平带电压主要受三方面因素的影响。第一个方面是功函数差, 第二个方面是绝缘体中的电荷, 第三个方面是在硅-二氧化硅界面的表面态。

当化学成分不同的导电材料连接在一起时, 它们的费米能级必须相等。这使得两种材料的导带发生相对移动, 正像我们在 p-n 结二极管的情形中看到的那样。

按照惯例, 材料的化学势是以真空为参考来测量的。这样量出来的势能称之为功函数, 它是将材料中费米能级上的电子移出到无穷远处所需的能量。这种测量有一定的复杂性, 而且它对固体器件的设计只有有限的意义。不过化学势的另一特性可以使功函数发挥作用。化学势遵守基尔霍夫电压定律。这样, 金属与半导体之间的功函数差将为 $\phi_{MS} = \phi_M - \phi_S$, 其中 ϕ_M 为金属的功函数, ϕ_S 为半导体的功函数。设金属与氧化物-半导体连接, 再从半导体连回金属, 由于这一循环中势的总和为零, 我们有

$$0 = \phi_{MO} + \phi_{OS} + \phi_{SM} \quad (2.36)$$

或

$$\phi_{MS} = \phi_{MO} + \phi_{OS} \quad (2.37)$$

注意从某种意义上说, 金属与半导体之间的氧化物是无关紧要的。对 MOS 器件有几种情况。如果栅是传统的金属铝, 那么我们关心的就是铝与硅之间的化学势差。对于本征硅, 化学势差 ϕ_{MS} 为 $-0.61V$, 栅为金的情形为 $-0.11V$, 这是因为铝与金有 $0.5V$ 的功函数差 [Glaser77, p.92]。对于掺杂的硅, ϕ_{MS} 随 ϕ_F 改变而改变。硅向 n 型变化时, ϕ_{MS} 增加, 硅向 p 型变化时, ϕ_{MS} 减小。对于铝栅, 我们有

$$\phi_{MS} = -0.61 + \phi_F \quad (2.38)$$

对于简并掺杂的多晶硅栅, 其数字上有所不同, 但定性方面是完全一样的。我们假定, 对于多晶硅可以采用能带图。当我们了解到在多晶硅上可以做出漏电较大的二极管, 而且多晶硅的再结晶也是一个非常活跃的研究领域 [Gibson80], 这假定就更具合理性了。对于 n^+ 多晶硅栅, 我们有

$$\phi_{MS} = -0.55V + \phi_F \quad (2.39)$$

其中 $0.55V$ 约为 E_G/q 的一半。

1) 此处原文中为 qE_G 。——译者注

氧化物中的电荷也要影响平带电压。氧化物中的一层电荷 Q 将在栅与衬底中感应出相反的电荷,而这要由平带电压来抵消。对于这种情况,由泊松方程的解可求出

$$\Delta V_{FB} = -\frac{x}{T_{ox}} \frac{Q}{C_{ox}} \quad (2.40)$$

其中 x 为氧化层中 Q 与栅的距离。对于一般情况,我们有

$$\Delta V_{FB} = -\frac{1}{C_{ox}} \int_0^{T_{ox}} \frac{x}{T_{ox}} \rho(x) dx \quad (2.41)$$

其中 $\rho(x)$ 为单位面积电荷密度。平带电压移动与电荷的位置有关。如果离子是可动的,它们将在自建的及外加的电场作用下移动。这将改变 V_{FB} 从而改变器件的阈电压。典型的可动离子电荷 Q_m 包括碱金属离子 Na^+ 、 K^+ 和 Li^+ 。在MOS发展的早期,由于这一机理,钠离子沾污就成了MOS器件阈电压不稳定性的主要原因。在氧化物中可能存在的其它类型的电荷有称之为氧化层俘获电荷的 Q_{ox} 。氧化层俘获电荷可能由X射线或高能(热)电子产生。

所有的硅MOS结构在硅-二氧化硅界面附近都有一层电荷。这一界面的确切本质至今尚未完全弄清楚。通常认为硅转变为二氧化硅的过渡中,首先经过一个 SiO_x 分子层,其中 $1 < x < 2$,然后经过约10至40Å的应力 SiO_2 。在这样的“界面”中有两种类型的电荷。其中之一是位于界面30Å之内的固定电荷 $Q_f^{(1)}$ 。另一种是界面俘获电荷 Q_{it} 。这一电荷大致与硅表面键的数目成正比。例如,(100)材料的 Q_{it} 通常比(111)材料低。这些界面电荷相应于硅禁带中的施主或受主能态。 $N_{it} (= Q_{it}/q)$ 可以小至 $10^{10} cm^{-2}$ 。 Q_{it} 与 Q_f 都与工艺密切相关, $N_f + N_{it}$ 的典型值为 1.5×10^{11} 离子/ cm^2 。

令方程(2.40)中的 $x = T_{ox}$,可以计算出表面电荷对 V_{FB} 的贡献:

$$\Delta V_{FB} = -\frac{Q_f + Q_{it}}{C_{ox}} \quad (2.42)$$

把平带电压的三个组成部分加起来,我们得到

$$V_{FB} = \Phi_{MS} - \frac{Q_f^{(1)} + Q_{it}}{C_{ox}} - \frac{1}{C_{ox}} \int_0^{T_{ox}} \frac{x}{T_{ox}} \rho(x) dx \quad (2.43)$$

所有这些参数都与制造工艺有关,因此一旦工艺确定之后,对电路设计者重要的只是 V_{FB} 。

2.2 MOS晶体管中的一级效应

在MOS结构的相对两侧各加一导体可将MOS电容器变为晶体管,如图2.10所示。这两个导体就是源端和漏端。像电容的情形一样,晶体管的沟道有三个工作区,它们是:

积累: $v_G < V_{FB}$ [图2.11(a)], 表面大量积累着衬底的多数载流子。在源和漏端有反向偏置的p-n结,源漏之间基本上没有电流。

耗尽: $v_G > V_{FB}$ [图2.11(b)], 在表面,多数载流子耗尽。沟道区既无载流子也无电流。在这一工作区,栅电容下降。

反型: $v_G \gg V_{FB}$, [图2.11(c)], 在表面有反型(少数)载流子,它们形成了源漏之

1) Q_f 以前写为 Q_{it} 。

间的导电通道。栅电压越高，载流子越多，电流也越大。

从电路设计的观点看，每一个工作区都是重要的。最重要的参数是阈电压，它将耗尽区与反型区（截止与导通）分开。为简单起见，我们将首先讨论源和衬底处在同一电位的情形。

阈电压

我们已经知道，在 MOS 结构上加一个正的 v_G ，可以在栅下感应出反型层和耗尽层。现在我们再一次考查这一结构来算出每一工作区的电荷。

定义 Q_s 为在硅中感应的单位面积总电荷， Q_G 为栅电极上的单位面积总电荷。电中性要求

$$Q_G + Q_s = 0 \quad (2.44)$$

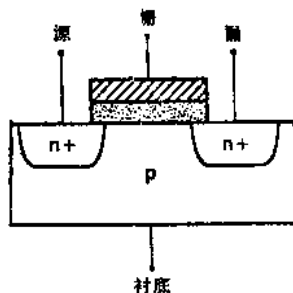


图 2.10 MOS 晶体管

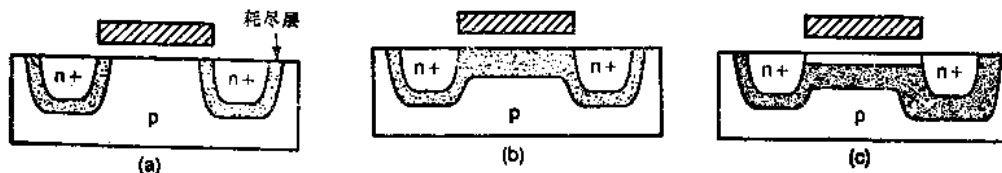


图 2.11 MOS 晶体管的三个工作区。(a) 积累；(b) 耗尽；(c) 反型

硅表面的电位为 ϕ_s 。我们可写出栅氧化层电容上的电荷与电压的关系：

$$Q_G = (v_G - V_{FB} - \phi_s) C_{OX} \quad (2.45)$$

注意当 $v_G = V_{FB}$ 时， $\phi_s = 0$ 因而 $Q_G = 0$ 。我们可以将硅中的电荷分为两个分量， Q_B 为耗尽区中的固定电荷， Q_N 为反型层中的可动电荷：

$$Q_s = Q_B + Q_N \quad (2.46)$$

利用方程 (2.45)，可得

$$v_G - V_{FB} = \phi_s - \frac{Q_B + Q_N}{C_{OX}} \quad (2.47)$$

图 2.12 所示为 MOS 晶体管中的各种电荷与电压， Q_N 与表面势 ϕ_s 的关系示于图 2.13 [Sze 81, p. 369]。

现在我们可以推导 MOS 晶体管阈电压的方程。当 MOS 晶体管进入反型区时开始导通。在反型区的边界， $Q_N = 0$ 。我们定义开始形成沟道时的栅电压为阈电压 V_{TX} 。根据方程 (2.47)，我们有

$$V_{TX} = -\frac{Q_B}{C_{OX}} + \phi_s + V_{FB} \quad (2.48)$$

硅的表面势必须达到能维持沟道中的电子才发生反型。在分析 MOS 电容时，我们论述了反型约发生于表面势为 $2\Phi_{FP}$ 时。这一分析假定是突然过渡到导通状态的。实际上过渡是指数性的，超过约 1/3 伏才达到强反型。尽管如此， V_{TX} 仍然是描述 MOS 晶体管特性的一个极其有用的参数，特别是当它们被用于数字电路的时候。在反型时， $\phi_s = 2\Phi_{FP}$ ， Q_B 记为 Q_{B0} ，因有

$$V_{TX} = -\frac{Q_{B0}}{C_{OX}} + 2\Phi_{FB} + V_{FB} \quad (2.49)$$

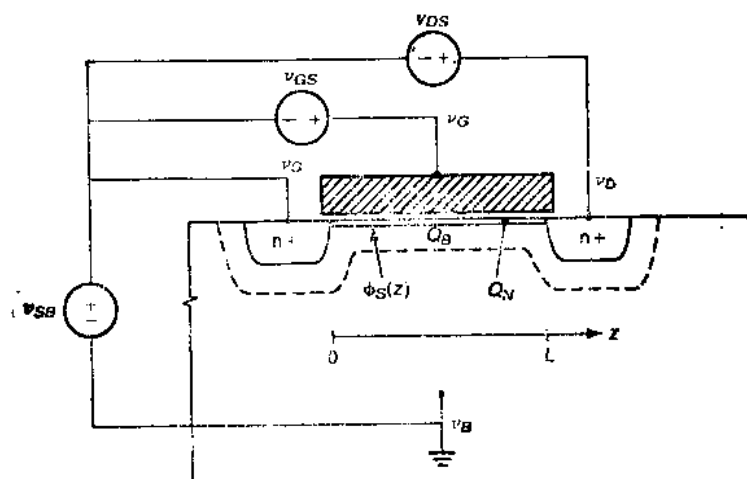


图 2.12 MOS 晶体管的电位与电荷

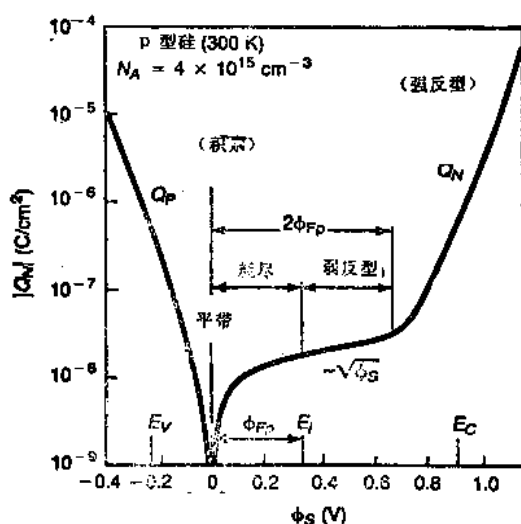


图 2.13 室温下, $N_A = 4 \times 10^{15} \text{ cm}^{-3}$ 时硅中电荷密度随表面势 ϕ_S 的变化 [Sze 81, p. 369]

耗尽区电荷 Q_B 公式可在对 p-n 结二极管的分析中找到。

温度升高时 V_{TX} 下降。对硅栅工艺, 温度系数典型值在 -1 到 $-3 \text{ mV}/^\circ\text{C}$ 之间。温度效应主要来源于 n_i 随温度的变化, 每升高 11°C 它增加一倍 [Sze 81, p. 20]。对 p 沟道晶体管, 温度升高, 阈值变正, 变化的速率与 n 沟相同。

电压控制的电阻

MOS 晶体管的源漏电流取决于沟道中可供导电的电荷量 Q_N 及电荷输运关系。

MOS 晶体管沟道中的自由电荷受到源漏电场的推动, 这一电荷的平均速度为

$$v = \mu \mathcal{E} \quad (2.50)$$

其中 $\mathcal{E}$ 为电场, μ 为载流子迁移率。迁移率决定于载流子有效质量、载流子与晶格原子的碰撞 (声子散射)、电荷中心 (氧化物电荷、界面电荷、杂质原子电荷) 的库仑散射、

氧化物界面的表面不平整度等。这些因素对 μ 的影响与外加的和自建电压以及制造方法有关。其中的物理机理是十分复杂的，大多数模型都带有经验性。后面我们将较详细地讨论迁移率。现在我们暂假定迁移率为常数。

为了得到一个物理概念，我们先分析 MOS 晶体管的线性工作区，在这一区中假定所有的电场和电荷分布都是均匀的。图1.6给出了 MOS 晶体管的截面和顶视图。假定源与衬底都接地。我们以漏与源之间的距离为沟道长度 L 。沟道宽度 W 示于顶视图。设对沟道深度积分的表面电荷密度 n 均匀，则电流为

$$i_D = qnWv = -\mu Q_N W \mathcal{E} \quad (2.51)$$

这里假定电场是由漏指向源的。

对均匀的横向电场，有

$$\mathcal{E} = -\frac{v_D}{L} \quad (2.52)$$

$$i_D = -\mu Q_N \frac{W}{L} v_D \quad (2.53)$$

其中 v_D 为漏对源的电压。微分电导 g_D 由下式给出：

$$g_D = \frac{\partial i_D}{\partial v_D} = -\mu \frac{W}{L} Q_N = -\mu S Q_N \quad (2.54)$$

其中 S 为方程(1.3)中定义的形状因子。根据方程(2.47)，有

$$Q_N = -\left(v_G - V_{FB} + \frac{Q_b}{C_{ox}} - \phi_s\right) C_{ox} \quad (2.55)$$

若假定 v_D 很小，则从源到漏的电位变化很小，因此有

$$\phi_s \approx 2\Phi_{FP} \quad (2.56)$$

由此，方程(2.55)成为

$$Q_N = -\left(v_G - V_{FB} - 2\Phi_{FP} + \frac{Q_{b0}}{C_{ox}}\right) C_{ox} \quad (2.57)$$

将方程(2.49)中的 V_{TX} 代入，则得

$$Q_N = -(v_G - V_{TX}) C_{ox} \quad (2.58)$$

由方程(2.54)可得

$$g_D = \mu S C_{ox} (v_G - V_{TX}) \quad (2.59)$$

对于 $v_D \ll v_G - V_{TX}$ ，电流由下式给出：

$$i_D = \mu S C_{ox} (v_G - V_{TX}) v_D \quad (2.60)$$

方程(2.60)只在漏电压小时才成立。下面我们将减少这些限制条件，采用同样的办法来推导有用的结果。

若 v_D 并不小，则 ϕ_s 的变化不能被忽略。至于 Q_b ，在这一分析中我们仍假定它沿沟道为常数。实际表明，这一假定对当今的 nMOS 电路是适用的。设

$$\phi_s(z) = 2\Phi_{FP} + \phi_c(z) \quad (2.61)$$

其中 z 为从源开始沿沟道的距离， ϕ_c 为沿沟道的电压。从方程(2.55)可得

$$Q_N(z) = -\left(v_G - V_{FB} - 2\Phi_{FP} + \frac{Q_{b0}}{C_{ox}} - \phi_c(z)\right) C_{ox} \quad (2.62)$$

此式可化简为

$$Q_N(\phi_c) = -(v_G - V_{TX} - \phi_c) C_{OX} \quad (2.62)$$

沟道的每一段都有一定的电阻。通以一定的电流 i_D ，在一段沟道上的压降为

$$d\phi_c = i_D dr \quad (2.64)$$

参看图 2.14，利用方程 (2.54) 可得

$$dr = dg_D^{-1} = -\frac{dz}{W\mu Q_N} \quad (2.65)$$

整理方程 (2.65) 并从源到漏积分，可得

$$\int_0^{v_D} W\mu Q_N d\phi_c = -i_D \int_0^L dz \quad (2.66)$$

或

$$i_D = -\mu S \int_0^{v_D} Q_N d\phi_c \quad (2.67)$$

代入 Q_N 值可得

$$i_D = -\mu S C_{OX} \int_0^{v_D} -(v_G - V_{TX} - \phi_c) d\phi_c$$

或

$$i_D = \mu S C_{OX} \left[(v_G - V_{TX}) v_D - \frac{1}{2} v_D^2 \right] \quad (2.68)$$

请注意它与前面方程 (2.60) 结果的相似性。除了 $v_D^2/2$ 项之外，两个方程是一样的，而该项对于很小的 v_D 接近于零。

图 2.15 所示为一定栅压下由方程 (2.68) 给出的 I_D - V_D 关系。如图中实线所示，电流起初随 V_D 增加而增加，达到峰值后开始下降。这一下降是不合理的，并且在实际器件中也没有观察到。测得的电流在峰值前与方程符合很好，此后进一步增加 V_D ，电流就饱和了（如虚线所示）。显然，在我们的分析中有些因素被忽略了。

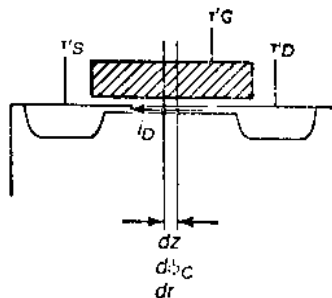


图 2.14 MOS 晶体管沿沟道的微分示意图

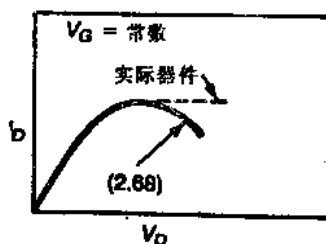


图 2.15 方程 (2.68) 中 I_D 与 V_D 的关系。超过峰值点后，方程预言的特性是不符合物理概念的

仔细考察方程 (2.63)，我们看到当 $v_G - V_{TX} = \phi_c$ 时 $Q_N(\phi_c)$ 为零。 ϕ_c 的最大值发生在沟道的漏端，该处 $v_G - V_{TX} = v_D$ 。在这一点，沟道消失了，方程 (2.62) 不再成立。

利用方程 (2.68)，取 i_D 对 v_D 的微分并令其为零，可算出电流的峰值，在这一峰值点，我们怀疑我们的模型可能不再成立。我们有

$$-\frac{\partial i_D}{\partial v_D} = \mu S C_{ox}(v_G - V_{TX} - v_D) = 0 \quad (2.69)$$

定义电流饱和时的电压为 V_{Dsat} , 我们有

$$V_{Dsat} = v_G - V_{TX} \quad (2.70)$$

在这一电压下, 沟道漏端的 $Q_N = 0$ 。这样, 问题就清楚了, 这时我们必须找出一个新的物理模型来描述当在漏端沟道消失时发生的情况。我们知道在这一区域会存在一个耗尽区, 并且自由电荷将被沟道电场推到耗尽区边沿。当它们到达耗尽区后, 它们将被沟道端点与漏之间的电场扫到漏端。高于 V_{Dsat} 的漏电压将加到耗尽区上而不会出现在反型区, 反型区的电压将限制为 V_{Dsat} 。这样, 电流将达到饱和, 饱和值决定于 V_{Dsat} 。与反向偏置 p-n 结二极管的耗尽区情况不同, 在沟道的漏端并不存在势垒, 电场的方向是使电荷加速而不是排斥它们。我们定义饱和电流为

$$I_{Dsat} \equiv \frac{\mu S C_{ox}}{2} (v_G - V_{TX})^2 \quad (2.71)$$

在实际器件中, 在 V_{Dsat} 以上, i_D 与 v_D 仍有一些关系。为模拟这一依赖关系, 必须考虑几个二级效应。在 2.3 节中我们将处理这些效应。

上面的论述都是对 $v_s = 0$ 的 nMOS 器件而言的, 与此同时, 我们也应了解 $v_s \neq 0$ 的情况。下面我们就考虑源对衬底偏置的效应。考虑图 2.12 中四个端子都独立的情况。我们将以 v_B 为基准 (即 $V_B \equiv 0$)。在沟道的源端, 电位为 $2\Phi_{FP} + v_s$ 。从方程 (2.67) 的积分限中我们可以直观地看到它的影响。不过, 这一电位在计算 Q_B 和 Q_N 时也有间接的影响。在沟道源端 Q_N 的新方程变为:

$$Q_N = -\left(v_G - V_{FB} - 2\Phi_{FP} + \frac{Q_B}{C_{ox}}\right)C_{ox} \quad (2.72)$$

定义 V_{TB} 为

$$V_{TB} \equiv V_{FB} + 2\Phi_{FP} - \frac{Q_B}{C_{ox}} \quad (2.73)$$

我们用单边突变结的公式来计算 Q_B :

$$Q_B = -\sqrt{2\epsilon_{si}qN_A(v_s + 2\Phi_{FP})} \quad (2.74)$$

则阈电压为

$$V_{TB} = V_{FB} + 2\Phi_{FP} + \frac{\sqrt{2\epsilon_{si}qN_A(v_s + 2\Phi_{FP})}}{C_{ox}} \quad (2.75)$$

我们定义体效应或背栅效应因子 γ 为:

$$\gamma \equiv \frac{\sqrt{2\epsilon_{si}qN_A}}{C_{ox}} \quad (2.76)^*$$

注意 Q_B 与 γ 之间的紧密关系。重新将阈电压写为 γ 的函数, 有

$$V_{TX} = V_{FB} + 2\Phi_{FP} + \gamma\sqrt{2\Phi_{FP}} \quad (2.77)^*$$

和

$$V_{TB} = V_{TX} + \gamma(\sqrt{v_s + 2\Phi_{FP}} - \sqrt{2\Phi_{FP}}) \quad (2.78)^*$$

体效应因子 γ 决定了一个器件的阈电压随源对衬底电压的漂移。CMOS 电路中的 γ 通常大于 nMOS 电路中的 γ , 这是因为大多数 nMOS 电路采用 -2 至 -3V 的衬底偏压

V_{BS} , 而在 CMOS 电路中 nMOS 器件的 V_{BS} 通常为零。这使得 nMOS 情形下衬底掺杂可以低一些¹⁾。从方程 (2.76) 我们看出降低衬底掺杂可以减小体效应。体效应参数在相当大程度上与工艺技术有关, 用老的技术 γ 值就比较大。我们可能碰到从小于 $0.1V^{1/2}$ 到超过 $1V^{1/2}$ 的 γ 值。

为推导 i_D , 我们按新的积分限对 Q_N 积分:

$$i_D = -\mu S \int_{v_S}^{v_D} Q_N(\phi_C) d\phi_C \quad (2.79)$$

代入 Q_N , 可得

$$i_D = -\mu S C_{OX} \int_{v_S}^{v_D} -(v_G - V_{TB} - \phi_C) d\phi_C \quad (2.80)$$

由此求出

$$i_D = \mu S C_{OX} \left[(v_{GS} - V_{TE}) v_{DS} - \frac{v_{DS}^2}{2} \right] \quad (2.81)$$

其中

$$v_{TK} \equiv v_D - v_S \quad (2.82)$$

$$v_{GS} \equiv v_G - v_S \quad (2.83)$$

跨导参数定义为²⁾

$$K' \equiv \frac{\mu C_{OX}}{2} \quad (2.84)^*$$

下面是对这三个工作区的结果的小结

截止区: $(v_{GS} - V_{TE}) \leq 0$

$$i_D = 0 \quad (2.85)^*$$

非饱和区: $(v_{GS} - V_{TE}) \geq v_{DS}$

$$i_D = K' S [2 (v_{GS} - V_{TE}) v_{DS} - v_{DS}^2] \quad (2.86)$$

饱和区: $(v_{GS} - V_{TE}) \leq v_{DS}$

$$i_D = K' S (v_{GS} - V_{TE})^2 \quad (2.87)$$

由于我们忽略了 Q_N 沿沟道的变化, 这些结果只有对 γ 低于 $0.2V^{1/2}$ 的工艺才是准确的。

沟道区中的电荷分布

利用上面已经建立的方程很容易求出沿沟道的电荷分布, 这是非常重要的, 它有助于建立对 MOS 器件工作机理的直观了解。

定性地说, 我们知道, 在非饱和工作区 (见方程 2.86), 从源移动到漏, 沟道中载流子的数目是减少了, 这是因为 MOS 电容上栅对沟道的电压减小了。由于电流在源端和漏端是一样的, 载流子在漏端必须移动的快一些以补偿其数目的减少。为了求出定量的关系, 我们从方程 (2.86) 出发, 在沟道的任何一点 z ($0 \leq z \leq L$), 都可应用方程 (2.86) 得出

$$i_D = \frac{W}{z} K' [2 (v_{GS} - V_{TE}) \phi_{CS} - \phi_{DS}^2] \quad (2.88)$$

1) 限制因素为源漏穿透, 见 2.3 节。

2) 此处 K' 不同于 SPICE 中的 KP , $KP = 2K'$, (K' 也称为导电因子)。

其中 ϕ_{cs} 为 z 点对源的电压。联立方程 (2.86) 与 (2.88) 可得

$$\frac{z}{L} = \frac{2(v_{GS} - V_{TB})\phi_{cs} - \phi_{cs}^2}{2(v_{GS} - V_{TB})v_{DS} - v_{DS}^2} \quad (2.89)$$

由此可解出 ϕ_{cs}

$$\phi_{cs} = v_{GS} - V_{TB} - \sqrt{(v_{GS} - V_{TB})^2 - [2(v_{GS} - V_{TB})v_{DS} - v_{DS}^2] \frac{z}{L}} \quad (2.90)$$

定义饱和比 X 为

$$X = \frac{v_{DS}}{v_{GS} - V_{TB}} \quad (2.91)$$

可将方程 (2.90) 改写为

$$\phi_{cs} = (v_{GS} - V_{TB}) \left[1 - \sqrt{1 - (2X - X^2) \frac{z}{L}} \right] \quad (2.92)$$

图 2.16 画出了不同饱和比下 ϕ_{cs} 与 z 的关系。在 z 点曲线的斜率代表沟道中该点的电场强度。

将 ϕ_{cs} 对 z 微分由此计算电场 $\mathcal{E}$ ，即

$$\mathcal{E} = -\frac{d\phi_{cs}}{dz} \quad (2.93)$$

可得

$$\mathcal{E} = -\frac{(v_{GS} - V_{TB})}{2L} \times \frac{(2X - X^2)}{\sqrt{1 - (2X - X^2) \frac{z}{L}}} \quad (2.94)$$

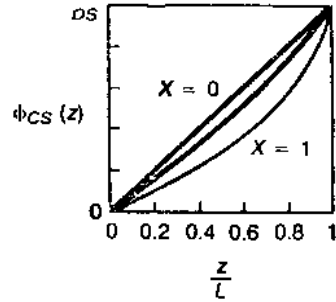


图 2.16 ϕ_{cs} 与 z/L 的关系

或写为

$$\mathcal{E} = -\frac{v_{DS}}{L} \frac{1 - \frac{X}{2}}{\sqrt{1 - (2X - X^2) \frac{z}{L}}} \quad (2.95)$$

图 2.17 所示为以 X 为参数，电场沿沟道的分布。对 $X=0$ ，电场为常数，其值为 v_{DS}/L 。对 $X=1$ ，在 $z=L$ 的位置电场趋向无穷。这是由于在我们的物理模型中，在这一点 Q_N 趋于零的结果。如果在我们的数学模型中把耗尽区也合在一起考虑将使进一步的计算非常困难。我们将只是提请注意在饱和区边界，我们的数学模型不是很准确的。

现在我们来考察电荷的定量分布。沟道中任一点的电荷为

$$Q_N = -(v_{GS} - V_{TB} - \phi_{cs})C_{OX} \quad (2.96)$$

在源端 ($z=0$) 有

$$Q_{source} = -(v_{GS} - V_{TB})C_{OX} \quad (2.97)$$

一般情况下有

$$\frac{Q_{channel}}{Q_{source}} = \frac{v_{GS} - V_{TB} - \phi_{cs}}{v_{GS} - V_{TB}} \quad (2.98)$$

其中 $Q_{channel}$ 为沟道中某点的电荷。通过饱和比方程 (2.98) 也可重写为

$$\frac{Q_{\text{channel}}}{Q_{\text{source}}} = \sqrt{1 - (2X - X^2) \frac{z}{L}} \quad (2.99)$$

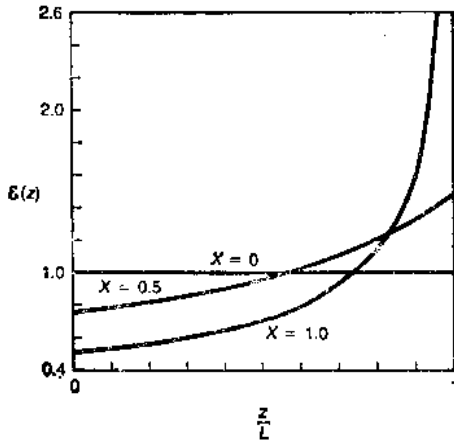


图 2.17 用 V_{DS}/L 归一化的 $\phi(z)$ 与 z/L 的关系

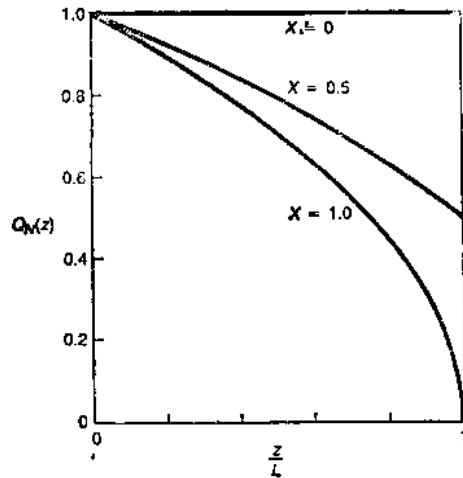


图 2.18 以源端值归一化的沟道可动电荷与 z/L 的函数关系

这一关系示于图 2.18。仔细研究图 2.16 至图 2.18 有助于了解 MOS 器件的工作特性。

漏电流的改进模型

由于 CMOS 工艺中通常衬底掺杂较高，因而 γ 和 Q_b 也较高。这使得 Q_b 的估算误差更大。下面我们将对 CMOS 工艺改进我们的模型，推导出对 Q_b 的较好近似。

方程 (2.85) 到方程 (2.87) 的推导中假定沟道下的耗尽区电荷 Q_b 为常数。由于从源到漏沟道的电位是变化的，这一假定在物理上是不真实的。误差的大小取决于 Q_b/Q_N 。当衬底重掺杂或栅氧化层比较厚时误差最大。现在我们在考虑 Q_b 变化的情况下重新推导电流方程。我们不是取 Q_b 的准确值而是把 Q_b 随 v_s 的变化近似地用泰勒展开。这样做是因为精确的方程很难处理，而且结果也并不准确很多。

由方程 (2.62)，我们得到 Q_N 的基本表达式，这里并未作简化。将这一方程应用于突变结，我们得到

$$\frac{Q_b}{C_{\text{ox}}} = \frac{-1}{C_{\text{ox}}} \sqrt{2e_{\text{si}} q N_A (v_s + \phi_{cs} + 2\phi_{fp})} \quad (2.100)$$

或写为

$$\frac{Q_b}{C_{\text{ox}}} = -\gamma \sqrt{v_s + \phi_{cs} + 2\phi_{fp}} \quad (2.101)$$

对沟道电压 ϕ_{cs} 作泰勒展开，取前两项，可得

$$Q_b \approx Q_b(v_s) + \phi_{cs} \left. \frac{dQ_b}{d\phi_{cs}} \right|_{\phi_{cs}=0} \quad (2.102)$$

此式可进一步化为

$$\frac{Q_b}{C_{\text{ox}}} = -\gamma \sqrt{v_s + 2\phi_{fp}} - \phi_{cs} \frac{\gamma}{2\sqrt{v_s + 2\phi_{fp}}} \quad (2.103)$$

代入方程 (2.55), 并用 V_{TE} 的方程 (2.78) 可得

$$Q_N = -(v_G - V_{TE} - \frac{\gamma \phi_{cs}}{2\sqrt{v_s + 2\Phi_{fp}}} - \phi_{cs} - v_s) C_{ox} \quad (2.104)$$

像前面一样, 为了求电流, 我们将电荷从源到漏对 ϕ_{cs} 积分:

$$i_{DS} = -\mu S \int_0^{v_{Dsat}} Q_N(\phi_{cs}) d\phi_{cs} \quad (2.105)$$

或者说

$$i_{DS} = \mu S C_{ox} \int_0^{v_{Dsat}} [v_{GS} - V_{TE} - (1 + \delta) \phi_{cs}] d\phi_{cs} \quad (2.106)$$

其中 δ 定义为

$$\delta = \frac{\gamma}{2\sqrt{v_{sb} + 2\Phi_{fp}}} \quad (2.107)^*$$

积分结果为

$$i_{DS} = K' S [2(v_{GS} - V_{TE})v_{Dsat} - (1 + \delta)v_{Dsat}^2] \quad (2.108)^*$$

这一方程与方程 (2.86) 是很相似的, 只是加上了一个修正因子 δ , δ 的作用是在较高的漏电压下使电流减少。沟道下的耗尽区在漏端最宽, 但在早先的计算中我们用的是源端的值, 漏电压越大, Q_s 的误差也越大。这意味着我们对 Q_s 值估计偏低, 因而对 Q_N 和电流估计偏高¹⁾。校正因子 δ 在衬底掺杂浓度较高、栅氧化层较厚的工艺中影响最大。

为了求饱和电压, 我们用一般方法求方程 (2.108) 的峰值, 即

$$\frac{di_{DS}}{dv_{DS}} = 0 = v_{GS} - V_{TE} - (1 + \delta)V_{Dsat} \quad (2.109)$$

由此可求出

$$V_{Dsat} = \frac{v_{GS} - V_{TE}}{1 + \delta} \quad (2.110)$$

下面是以上关于三个工作区的分析结果的一个小结:

截止区: $(v_{GS} - V_{TE}) \leq 0$

$$i_{DS} = 0 \quad (2.85)^*$$

非饱和区: $(v_{GS} - V_{TE}) \geq (1 + \delta)v_{DS}$

$$i_{DS} = K' S [2(v_{GS} - V_{TE})v_{DS} - (1 + \delta)v_{DS}^2] \quad (2.108)^*$$

饱和区: $(v_{GS} - V_{TE}) \leq (1 + \delta)v_{DS}$

$$i_{DS} = K' S \frac{(v_{GS} - V_{TE})^2}{1 + \delta} \quad (2.111)^*$$

其中

$$K' \equiv \mu \frac{C_{ox}}{2} \quad (2.84)^*$$

$$V_{TE} = V_{TX} + \gamma (\sqrt{v_{sb} + 2\Phi_{fp}} - \sqrt{2\Phi_{fp}}) \quad (2.78)^*$$

$$V_{TX} = V_{FB} + 2\Phi_{fp} + \gamma \sqrt{2\Phi_{fp}} \quad (2.77)^*$$

1) 尽管我们用泰勒展开推导出 δ , 泰勒级数的前两项并不一定是函数的最佳近似。在这一情形, 实际上 δ 太大, 不可能提供对 $Q_N(z)$ 的最佳线性拟合。由于这一原因, 有时采用比方程 (2.107) 所给出的 δ 略小一些的 δ 值。

$$\gamma \equiv \frac{\sqrt{2\epsilon_{si}qN_A}}{C_{ox}} \quad (2.76)^*$$

$$\delta \equiv \frac{\gamma}{2\sqrt{v_{SB} + 2\phi_{fp}}} \quad (2.107)^*$$

这些方程是本节的主要结果。本书中它们将全部被用于手算。

例 2.3 MOS 器件参数计算。

MOS 器件衬底为 P 型、电阻率为 $20\Omega\cdot\text{cm}$ 、晶向为 (100)。栅材料为简并掺杂的 n^+ 多晶，栅氧化层厚度为 $300\text{\AA}$ 。

从图 2.19 可以查出 $20\Omega\cdot\text{cm}$ 材料相应于掺杂浓度 $N_A = 9 \times 10^{14}\text{cm}^{-3}$ [Irvin 62]。由方程 (2.11) 可求出 $\phi_{fp} = -0.285\text{V}$ ，或 $2\phi_{fp} = 0.570\text{V}$ 。

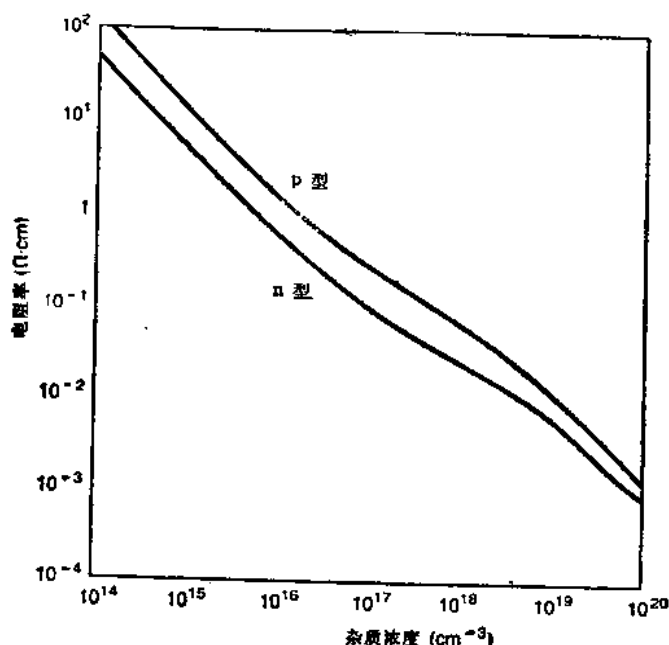


图 2.19 300K 下硅电阻率与施主及受主杂质浓度的关系 [Irvin 62]

由方程 (2.28) 我们还可求出 $C_{ox} = 0.00115\text{F}/\text{m}^2 = 0.00115\text{pF}/\mu\text{m}^2$ 。反型时的 X_{pmax} 可由方程 (2.34) 给出，由此求出 $X_{pmax} = 0.91\mu\text{m}$ 。反型时，半导体的 Q_s 为 $Q_{smax} = -qN_A X_{pmax} = -1.3 \times 10^{-4}\text{C}/\text{m}^2 = 814$ 个离子/ μm^2 ，其电容为

$$C_{si} = \frac{e_{si}}{X_{pmax}} = 1.14 \times 10^{-4} \quad (\text{pF}/\mu\text{m}^2)$$

可求出 $C_{min} = 1.04 \times 10^{-4}\text{pF}/\mu\text{m}^2$ ， $C_{max}/C_{min} = 11$ 。注意，按照 Goetzberger 的准确计算，实际上 C_{max}/C_{min} 比 10 小一些 [Goetzberger 66, p.1121]。我们的误差来自 $\phi_s = 2\phi_{fp}$ 的近似。

利用 $N_f = 1.5 \times 10^{11}\text{cm}^{-2}$ 及 $Q_{if} = \rho(x) = 0$ 来计算 V_{FB} 。根据方程 (2.39)， $\Phi_{MS} = -0.835\text{V}$ 。由方程 (2.43) 可得

$$V_{FB} = \Phi_{MS} - \frac{qN_f}{C_{ox}} = -1.04 \quad (\text{V})$$

由方程 (2.49) 我们计算出 $V_{Tx} = -0.37V$, 由方程 (2.76) 算出 $\gamma = 0.15V^{1/2}$, 并由方程 (2.107) 算出 $\delta = 0.1$.

MOS 晶体管的电容特性

尽管我们对四端 MOS 晶体管电流-电压特性的分析在代数学上有点复杂, 但却是建立在稳固的理论基础上的。由于在正常情况下直流电流只在源漏之间流通, 相对说问题比较简单。而另一方面, 电容则存在于任意两个端子间。[Meyer 7]

我们暂且只考虑微分电容 (见附录 B)。我们可以把晶体管特性在工作点附近作线性近似, 这样我们就可以应用阻抗的概念。以 MOS 晶体管的一个端子作为基准, MOS 晶体管的电容矩阵就包含九个独立分量。例如, 它们包括 C_{GS} 和 C_{SD} 。在我们熟悉的大多数情况下, 这两个电容是相同的, 也就是说, 电容矩阵是互易的。这里有一些棘手的理论问题。例如, 人们知道, 若网络具有非互易电容矩阵, 在无穷大频率下它可以产生无穷大功率。另一方面, 人们还知道, 对 MOS 晶体管电容特性的最佳拟合会导致非互易的电容。这些非互易电容已经被测出来了 [Paulos 82]。问题之一在于晶体管为一分布结构, 如果我们要求十分严格, 就必须采用传输线来模拟 [Paulos 83]。显然, 这是太复杂太精确了, 完全超出了我们的需要, 所以我们删去了这些高次的效应, 并且得到了互易的电容。在本书中, 我们将避开所有这些问题, 只讨论最简单的模型。不过我们应该清醒: “冰层是薄的”。

当 MOS 晶体管处于非截止状态下, 从概念上我们可以把栅电容分解为栅源分量和栅漏分量 (如图 2.20 所示)。 C_{GS} 及 C_{GD} 定义为

$$C_{GS} = \left. \frac{\partial Q_G}{\partial v_{GS}} \right|_{v_{GD} = \text{常数}} \quad (2.112)$$

$$C_{GD} = \left. \frac{\partial Q_G}{\partial v_{GD}} \right|_{v_{GS} = \text{常数}} \quad (2.113)$$

利用方程 (2.63) 我们可以求出沟道中任一点的电荷密度 Q_N 。根据方程 (2.46), $Q_s = Q_N + Q_b$, 其中 Q_s 为硅中感应的电荷总密度。 Q_s 必须与等量的栅电荷相平衡, 因而 $Q_G = -Q_s = -(Q_N + Q_b)$ 。利用方程 (2.63) 及方程 (2.99), 我们得到

$$Q_N = -(v_{GS} - V_{TE}) C_{OX} \sqrt{1 - (2X - X^2) \frac{z}{L}} \quad (2.114)$$

其中 X 为饱和比, 方程 (2.114) 中忽略了体效应。

为了求出沟道的总电荷, 我们必须把 Q_N 乘以沟道宽度, 并沿沟道积分。即

$$Q_{TN} = -(v_{GS} - V_{TE}) W C_{OX} \int_0^L \sqrt{1 - (2X - X^2) \frac{z}{L}} dz \quad (2.115)$$

或

$$Q_{TN} = (v_{GS} - V_{TE}) C_{OX} W L \frac{2(X^2 - 3X + 3)}{3(X - 2)} \quad (2.116)$$

其中 Q_{TN} 为沟道中感应的总电荷。方程 (2.112) 和 (2.113) 可按间接微分规则通过

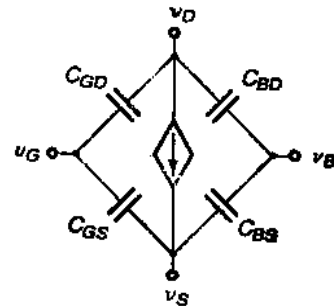


图 2.20 MOS 晶体管的简单电容模型

Q_{TN} 求出如下:

$$C_{GS} = -\frac{\partial Q_{TN}}{\partial v_{GS}} = -\frac{\partial Q_{TN}}{\partial X} \frac{\partial X}{\partial v_{GS}} \quad (2.117)$$

$$C_{GD} = -\frac{\partial Q_{TN}}{\partial X} \frac{\partial X}{\partial v_{GD}} \quad (2.118)$$

X 中的 $v_{DS} = v_{GS} - v_{GD}$ 。计算可得

$$C_{GS} = \frac{2}{3} W L C_{OX} \left[1 - \frac{(1-X)^2}{(2-X)^2} \right] \quad (2.119)$$

$$C_{GD} = \frac{2}{3} W L C_{OX} \left[1 - \frac{1}{(2-X)^2} \right] \quad (2.120)$$

对极端的 X 值, 有

$$C_{GS} = C_{GD} = \frac{1}{2} C_{OX} W L \quad X = 0 \quad (2.121)^*$$

$$C_{GS} = \frac{2}{3} C_{OX} W L \quad X = 1 \quad (2.122)^*$$

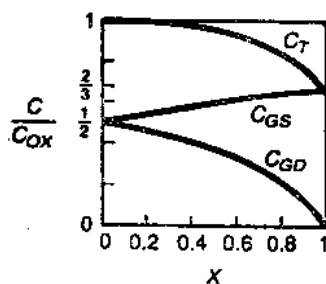
和

$$C_{GD} = 0 \quad (2.123)^*$$

这里 X 是由方程 (2.91) 定义的饱和比。对方程 (2.121) 有个简单、直观的解释: 由于 $X = 0$ 时源和漏没有区别, 因而电容就应该相等。当 $X = 1$ 时, 器件饱和, 其工作与漏电压无关, 所以方程 (2.123) 也是很直观的。关于方程 (2.122), 我们只好相信数学了。

图 2.21 示出了方程 (2.119) 和 (2.120)。图中也画出了 C_T , $C_T = C_{GS} + C_{GD}$ 。虽然 $X \geq 1$ 时我们的方程不成立, 不过我们很容易看出, 超过饱和点后, C_{GS} 和 C_{GD} 将和方程 (2.122) 和 (2.123) 给出的值一样。

这些方程对多数数字应用是足够了。不过偶尔我们也需要计算沟道与衬底之间的电容。像在栅对沟道电容的情形中一样, 我们把这一电容也分解为衬底对源的电容 C_{BS} 和衬底对漏的电容 C_{BD} 。在 C_{GS} 和 C_{GD} 的分析中, 我们假定 Q_B 为常数。这将导致 $C_{BS} = C_{BD} = 0$ 。实际上 Q_B 是变化的, 当我们分析体效应时我们看到了这一改变。在分析体效应时, 我们实际假定了 v_{DS} 为常数。在这一条件下, 我们有



$$\frac{1}{C_{OX}} \frac{dQ_B}{dv_s} = \frac{dV_{TB}}{dv_s} \Big|_{V_{DS} = \text{常数}} \quad (2.124)$$

或写为

$$\frac{dQ_B}{dv_s} = b C_{OX} \quad (2.125)$$

这里我们把 b 定义为:

$$b \equiv \frac{dV_{TB}}{dv_s} \quad (2.126)$$

分析表明 [Tsividis 80]

$$C_{BS} = b C_{GS} \quad (2.127)$$

$$C_{BD} = b C_{GD} \quad (2.128)$$

在电路模拟程序中, 实现非线性电容是相当困难的, 这是因为其数字误差影响很大。若程序并不显含地跟踪电荷, 积分误差会使得电路的电荷好像不守恒一样。这并不是模

型的问题，而是在模拟程序中如何选择主、次变量。在这一问题被理解前写出的大多数早期的模拟程序显然是做了错误的选择。这也包括 SPICE2.G。因此，动态 RAM、开关电容滤波器、电荷耦合器件等的模拟不能被指望以维持电荷，而且在这些专门应用中往往会导致很没有把握的结果。

例 2.4 $Q_B(z)$ 与饱和比的函数关系。

利用方程 (2.101) Q_B 的完全表达式和方程 (2.92) 中 $\phi_c(z)$ 的近似式，我们求出

$$\frac{Q_B(z)}{Q_B(0)} = \sqrt{1 + \frac{v_{GS} - V_{TE}}{v_{SB} + 2\phi_{FP}} \left[1 - \sqrt{1 - (2X - X^2) \frac{z}{L}} \right]}$$

其中 $Q_B(0)$ 为源端的体电荷。这一函数示于图 2.22。注意， $X = 1$ 时

$$\frac{Q_B(L)}{Q_B(0)} = \sqrt{1 + \frac{V_{DSAT}}{v_{SB} + 2\phi_{FP}}}$$

与推导用 δ 的电流表达式时一样，在同样条件下用 Q_B 的泰勒展开，可得

$$\frac{Q_B(L)}{Q_B(0)} = 1 + \frac{V_{DSAT}}{v_{SB} + 2\phi_{FP}}$$

显然，对于较大的饱和电压，所得结果也偏大。

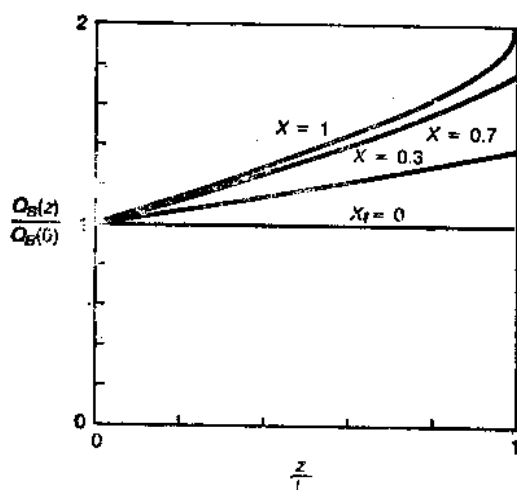


图 2.22 MOS 晶体中沿沟道的耗尽层电荷

表 2.1 阈值方程小结

n 沟	p 沟
$\phi_{FP} = -\frac{kT}{q} \ln \frac{N_A}{n_i}$	$\phi_{FN} = -\frac{kT}{q} \ln \frac{N_D}{n_i}$
$X_{Fmax} = \sqrt{\frac{2\epsilon_{Si}(2 \phi_{FP} + V_{SB})}{qN_A}}$	$X_{Nmax} = \sqrt{\frac{2\epsilon_{Si}(2 \phi_{FN} + V_{SB})}{qN_D}}$
$V_{TX} = V_{FR} + 2 \phi_{FP} + \gamma\sqrt{2 \phi_{FP} }$	$V_{TX} = V_{FR} - 2 \phi_{FN} - \gamma\sqrt{2 \phi_{FN} }$
$\gamma = \frac{\sqrt{2\epsilon_{Si}qN_A}}{C_{OX}}$	$\gamma = \frac{\sqrt{2\epsilon_{Si}qN_D}}{C_{OX}}$
$\phi_s(\text{反型}) = 2 \phi_{FP} + V_{SB}$	$\phi_s(\text{反型}) = -2 \phi_{FN} + V_{SB}$
$V_{TE} = V_{TX} + \gamma(\sqrt{V_{SB} + 2 \phi_{FP} } - \sqrt{2 \phi_{FP} })$	$V_{TE} = V_{TX} - \gamma(\sqrt{ V_{SB} + 2 \phi_{FN} } - \sqrt{2 \phi_{FN} })$
$\phi_{FP} < 0$	$\phi_{FN} > 0$
$V_{SB} \geq 0$	$V_{SB} \leq 0$

2.3 MOS 晶体管中的二级效应

在前几节我们发展了一个解析模型，通过少数基本物理参数来预言器件的特性。不幸的是，实际器件有一些附加的重要影响因素。上面我们所考察的这些效应用以模拟 1975 年左右的工艺是合用的。对于比较现代的工艺，上一节的方程所预言的电流与实际可能有因子为 2 的偏差。所以我们必须把这些附加效应包括在内才能把我们的模型应用于实际电路。这些效应，一部分可以采用解析模型，而另一些则需要用经验公式。

短沟和窄沟器件模型每天都有新发展，关于这些效应的更精确的公式肯定还会继续发表。本节中的材料对阅读进一步的著作会有教益。本节的主要内容涉及这些二级现象的物理机制及它们影响 MOS 器件外特性的一般途径。不管人们是否已完全了解清楚，这些现象确实影响着 VLSI 系统的性能以至功能。我们将从两个最重要的、也是最简单的效应出发。

长度和宽度的变化

当设计一个电路时，最直接受我们控制的两个参数就是器件宽度 W 和长度 L 。由于几方面的原因，这些参数的最后值与设计值并不精确吻合。一个最明显的原因是制造的公差。第四章将讨论所有设计参数制造公差的影响并提出最坏情况模型的概念。除了统计波动之外， W 和 L 值的偏移还有一些基本的物理原因。

先讨论器件宽度 W 的情形，定义

$$W \equiv W_{\text{drawn}} - 2\Delta W \quad (2.129)^*$$

ΔW 为描述宽度工艺变化的参数。我们假定这一变化独立于版图尺寸及其它基本工艺参数（这就是说，它是一个独立的工艺参数）。在这一定义下， ΔW 可以用于说明一些不同的效应。在现代的局部氧化、硅栅工艺中， ΔW 主要由两方面现象引起。

头一个现象是场氧化过程中（有源区）器件边界的物理移动。在局部氧化工艺中，用一个有源区掩模来确定防止场氧化的区域。通常这是靠在氧化前盖上一块氮化硅来实现的。这块氮化硅阻挡氧由表面流向硅。在氮化硅块的边缘则有斜对角方向的氧源。其

结果是通常所谓的“鸟嘴”，如图 2.23 及

1.5 节所示，其成因是明显的。鸟嘴的准确形状是氧化参数的函数，但不管是什么情况，它都使得晶体管的有效宽度下降。在简单的工艺中， ΔW 侵蚀与场氧化层厚度具有同一量级。

第二个现象是当用场注入来控制表面漏电时发生的。由于我们尚未讨论离子注入，我们需要稍稍离题考察一下有关的要点。

离子注入工艺提供了调整硅中杂质浓度的方便手段。离子注入机基本上是一个

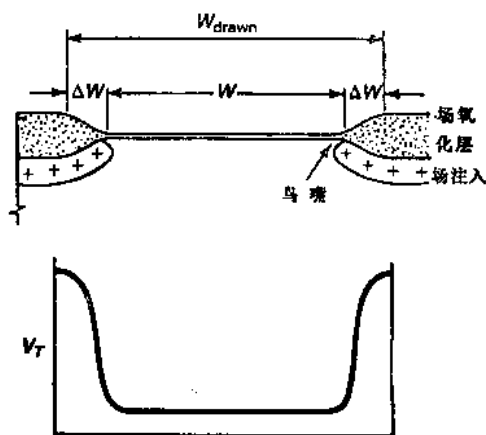


图 2.23 器件的宽度定义

杂质离子线性加速器。它包括一个离子源,一个加速管(以 30-300kV 范围内可调的静电场加速), 和一个安放硅片的靶区。调整加速电压可以精确控制注入深度。调整束流和注入时间可以精确控制注入离子杂质剂量。在不同的工艺阶段, 有几不同的掩蔽方法使注入杂质进入选定区域。

MOS 工艺中, 氧化层厚度是区别有源(晶体管)区和无源(场)区的主要参数。不过互连跨过厚场区也形成一个基本的 MOS 晶体管结构。器件是否有源取决于它的阈值。由方程 (2.75) 有

$$V_{TS} = V_{FB} + 2\Phi_{fp} + \frac{1}{C_{ox}} \sqrt{2\epsilon_{si}qN_A(v_s + 2\Phi_{fp})} \quad (2.130)$$

由于 C_{ox} 与氧化层厚度成反比, 厚场氧化层将使 V_{TS} 增加。增加的幅度也取决于衬底杂质浓度 N_A 。将厚氧化层与控制离子注入的 N_A 结合起来, 工艺工程师可以使 V_{TF} (厚场阈值) 远高于大多数电路的工作电压。

场注入通常在场氧化前进行, 它以氮化硅区作为注入的掩蔽。在高温的场氧化过程中, 杂质原子将由于浓度梯度而向硅中(以及场氧化层中)扩散。其结果如图 1.5 所示, 形成了一个侵入有源晶体管区的杂质浓度较高的掺杂带。由于场区 N_A 较大, 除了向有源正侵蚀之外, 它还使扩散 n^+ 线侧壁电容增加、击穿电压下降。因此要采取一定的折衷, 这就限制了 V_{TF} 的幅度。

由于氧化层厚度坡度和杂质浓度分布两方面的原因, 在晶体管边缘引起阈电压的连续变化, (如图 2.23 所示)。若此曲线足够陡, 我们可以把它近似为一个阶梯, 并用 ΔW 作为总的参数。由于 W_{drawn} ($W_{版图}$) 比 ΔW 大得多, 我们近似的误差属三级效应。不过注意在非常窄 ($W \leq 2\Delta W$) 的器件中, 情况就很不同了。

我们定义有效沟道长度 L 为:

$$L \equiv L_{drawn} - 2L_{diff} - 2\Delta L_{poly} \quad (2.131)^*$$

L_{diff} ($L_{扩散}$) 代表栅下源和漏的横向扩散 (一般为结深 X_j 的 0.7 倍)。多晶硅栅的 $L_{栅面}$ 与 L_{final} ($L_{栅底}$) 之差为 $\Delta L_{多晶}$ (L_{poly}):

$$\Delta L_{poly} = \frac{L_{drawn} - L_{final}}{2} \quad (2.132)$$

这些定义如图 2.24 所示。我们把 L 分为两个部分的原因是我们需要用这两个参数

来正确地计算栅电容。前面对 MOS 电容的讨论只包括晶体管的沟道区 (换句话说只是晶体管的本征部分)。实际晶体管的栅与源及栅与漏之间还有两个由薄氧化层与侧向扩散区覆盖引起的附加固定电容。这些电容可由下式计算:

$$C_{GSO} = WL_{diff}C_{ox} \quad (2.133)$$

$$C_{GDO} = WL_{diff}C_{ox} \quad (2.134)$$

其中 C_{GSO} 为栅源覆盖电容, C_{GDO} 为栅漏覆盖电容。在所有器件方程中的 W 和 L 都应采用经过修正的器件宽度和长度值。

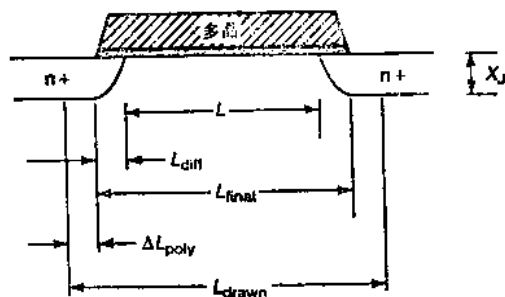


图 2.24 器件长度的定义

迁移率退化

关于手算模型我们就讨论到这里。在本节剩下部分我们将讨论的大部分模型主要用于计算机辅助设计(CAD)程序,虽然也有例外。不过这并不意味着我们不需要研究这些模型。这有几个基本的理由。首先,这些效应可能成为一个电路能否工作的原因。这并不是偶然情况,而是经常发生的。为了能够进行设计,就要懂得这些效应在什么地方以及如何发生重要作用。人们也必须理解CAD工具并不是完全精确的。即使是最好的CAD工具也有内在的应用条件。某些“创造性”的工作的一个料得到的结果是,要将程序应用于某些情形,而其特性是程序作者从未想像过的。我们如何知道程序干得对不对?首先,我们必须成为能够充分利用该工具的专家。其次,我们必须知道正确的答案是什么,或者有没有违反程序的假设。

当我们引入迁移率时,我们假定它是常数,同时我们也指出这是一个过于简单化的假设。虽然改变迁移率的物理机理有若干种,但从器件引出端看进去是三种效应。第一个是特征迁移率 μ_0 。它是制造工艺的函数,与界面电荷密度、衬底杂质浓度,及硅片表面晶向等有关。特征迁移率与器件温度有很大关系,而且是器件特性随温度变化的主要原因。它的退化有几方面因素。我们把这些因素分解为垂直及水平场迁移率退化,每个退化成份通常用一个或多个经验参数来表示。

载流子的速度饱和导致水平场迁移率退化。当载流子被加速到很大的水平速度时,将发生速度饱和。这一效应通常以测量到的饱和速度 v_{sat} 来表征。尽管n型硅的 μ_0 比p型硅的 μ_0 大得多,但二者的饱和速度是差不多的。在高性能MOS器件的情形,载流子在沟道的大部分是以饱和速度漂移的,在这种情形中,p沟晶体管与n沟晶体管特性的匹配情况比长沟器件好得多。这并不是由于p沟器件改进了,而是由于n沟器件退化了。

为了确定 μ_0 ,人们通常在低的外加垂直和水平电场下对器件进行测量。根据方程(2.108),若 W 和 L 已知,则与迁移率成比例的 K' 可以在不同的条件下来确定。为了确定 μ ,必须独立地求出 C_{ox} 。这可以通过电容测量或在工艺中氧化层厚度的测量来实现。

我们可以把 μ 表示为

$$\mu = \mu_0(T) f_v(v_G, v_S, v_D) f_h(v_G, v_S, v_D) \quad (2.135)$$

其中 f_v 为垂直场退化函数, f_h 为水平场退化函数, T 为温度。迁移率随温度的变化可由下述方程描述:

$$\mu_0(T_2) = \mu_0(T_1) \left(\frac{T_2}{T_1} \right)^{-M} \quad (2.136)$$

其中 T_1 和 T_2 为绝对温度, M 为经验常数¹⁾,通常取为1.5到2之间。室温下n沟MOS器件 μ_0 的典型值为 $600\text{cm}^2/\text{V}\cdot\text{s}$,p沟情形为 $250\text{cm}^2/\text{V}\cdot\text{s}$ 。

往往采用如下方程作为垂直场函数

1) SPICE程序中采用 $M=1.5$,这个值对体内是正确的,而根据Sabnis和Clemens[Sabnis 79]的报道,在增强型n沟MOS晶体管反型层中 $M=2$ 。我们的经验是 $M \approx 1.6$ 。

$$f_v = \begin{cases} 1 & \text{对 } v_v \leq V_c \\ (V_c/v_v)^v & \text{对 } v_v \geq V_c \end{cases} \quad (2.137)$$

其中垂直电压 v_v 取为栅与沟道之间的平均压降, 在方程 (2.137) 中

$$v_v = v_{GS} - V_{TB} - \frac{v_{DS}}{2} \quad (2.138)$$

V_c 为临界电压, 等于 $\mathcal{E}_c T_{ox}$ 。临界电场 $\mathcal{E}_c$ 数量级为 $2 \times 10^5 \text{ V/cm}$, U 对 n 沟器件近似为 0.25, 对 p 沟器件为 0.15。 n 沟 MOS 晶体管的 μ 测量值与平均垂直电场及温度的函数关系示于图 2.25 [Sabnis 79]。

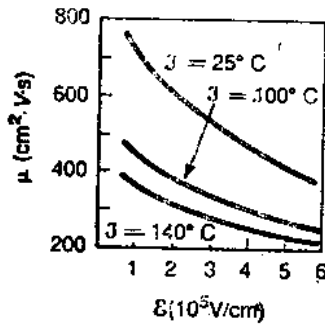


图 2.25 n 沟反型层的迁移率与垂直电场的关系 [Sabnis 79, p. 21]

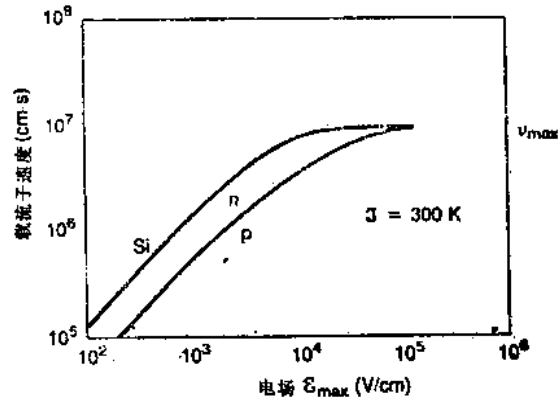


图 2.26 高阻硅载流子速度与横向电场的关系。饱和速度基本上与杂质浓度无关 [Smith 80, p. 797]

有时用简化的方程来表述 f_v (例如在著名的 SPICE2 电路模拟程序中), 其方程为

$$f_v = \frac{1}{1 + \theta (v_{GS} - V_{TB})} \quad (2.139)$$

其中 θ 为一经验常数。虽然通常 θ 是由测量确定的, 但对特定的工艺可以用方程 (2.137) 和方程 (2.138) 估计出来¹⁾。对典型的器件, 在大的栅压下, 由于垂直场退化, $I_{D,sat}$ 可能降低 25—50%。

横向电场迁移率退化比垂直电场退化对我们的器件电流方程有更大的作用。图 2.26 所示为 n 型和 p 型载流子速度与外加电场的关系。由图可见, 速度饱和的特征是有一个最大载流子速度 v_{max} , 或者说有一个临界电场 $\mathcal{E}_{max}$, 在该电场下速度达到饱和。电子饱和速度与温度的关系如图 2.27 所示 [Jacoboni 77]。

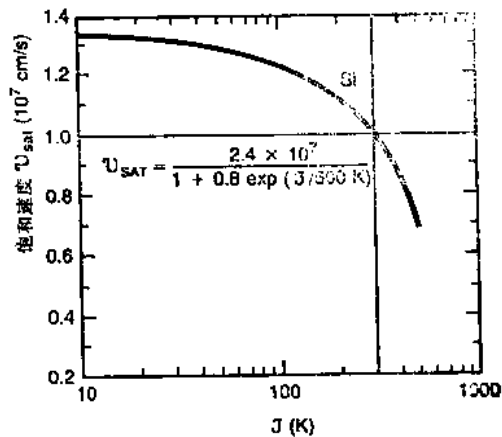


图 2.27 硅中电子饱和速度与温度的关系 [Jacoboni 77, p. 84]

在 2.2 节中我们看到, 在长沟模型基本假设下, 饱和点的横向电场为无穷。定性地

- 1) 基本上有两种曲线拟合法来求器件参数。一种办法是找出曲线的某一个区域, 在此区域中某一参数起主要作用并且从这一区域的局部数据提取该参数。用这种方法提取参数的一个显著例子是 V_T 。这样提取的参数又可作为已知参数用于提取下一个参数。一般说, 这样提取的参数具有明显的物理意义。另一种办法是同时考虑所有的参数和所有的数据, 并给出模型与数据的最佳拟合 [Ward 82]。这种办法提取的参数与数据拟合较好, 但在物理内涵上却有一定的牺牲。

说,可以预计,一旦载流子速度达到 $v_{\max}$,漏电压再进一步增加将对电场没有影响。从前面的分析我们知道最大电场发生在漏附近。为了计算饱和点,我们在漏附近应用连续方程,有

$$i_{DS} = -WQ_N(z=L)v_{\max} \quad (2.140)$$

利用方程 (2.108) 和 (2.140) 以及在漏点 $v_{DS} = \phi_{GS} = V_{Dsat}$ 的条件 (其中 V_{Dsat} 为考虑速度饱和后的新值), 我们得到

$$\begin{aligned} & \frac{\mu SC_{OX}}{2} [2(v_{GS} - V_{TE})V_{Dsat} - (1 + \delta)V_{Dsat}^2] \\ & = WC_{OX}(v_{GS} - V_{TE} - (1 + \delta)V_{Dsat})v_{\max} \end{aligned} \quad (2.141)$$

解 V_{Dsat} , 我们得到

$$V_{Dsat} = \frac{v_{GS} - V_{TE}}{1 + \delta} + \frac{v_{\max}L}{\mu} - \sqrt{\left(\frac{v_{GS} - V_{TE}}{1 + \delta}\right)^2 + \left(\frac{v_{\max}L}{\mu}\right)^2} \quad (2.142)$$

在方程 (2.141) 中应把垂直场迁移率退化考虑在内, 所以 μ 取为 $\mu_0 f_v$ 。注意当 $v_{\max}$ 趋于无穷时, 方程趋于原先的方程 (2.110)。方程 (2.142) 给出的 V_{Dsat} 值比方程 (2.110) 给出的小一些。为了说明迁移率的逐渐减小以致达到饱和点 V_{Dsat} , 我们写出水平场退化因子

$$f_h = \begin{cases} \frac{1}{1 + \mu v_{DS}/(v_{\max}L)} & \text{当 } v_{DS} \leq V_{Dsat} \\ \frac{1}{1 + \mu V_{Dsat}/(v_{\max}L)} & \text{当 } v_{DS} > V_{Dsat} \end{cases} \quad (2.143)$$

这里 μ 仍取为 $\mu_0 f_v$, 其中 μ_0 与温度的关系由方程 (2.136) 给出。对于沟道很短的器件, 在高电压下速度饱和具有决定性作用。由方程 (2.51), 有

$$i_{DS} = -Q_N W v \quad (2.144)$$

饱和区电流与 $v_{GS} - V_{TE}$ 的平方关系来自这一事实——当 $v_{GS} - V_{TE}$ 增加时 Q_N 与 v 二者都增加。当速度饱和时, v 近似等于 $v_{\max}$, 随 $v_{GS} - V_{TE}$ 改变的电流成分只有 Q_N 。所以, 对强场器件 $I_{DS} \propto v_{GS} - V_{TE}$ 而不是 $(v_{GS} - V_{TE})^2$ 。注意电流方程中也不再含有器件长度 L 。

短沟器件有一个与载流子速度有关的可靠性问题, 即热电子效应。热电子获得足够的能量就可以离开沟道区。例如, 电子可能越过硅-二氧化硅势垒并被氧化层陷阱所俘获, 这会引起长期的平带电压漂移 (通过 Q_{ot})。这一效应在低温时尤其严重, 因为这时陷阱退火的可能性很小。其它的热电子则参与衬底电流。为了减小热电子效应, 同时也由于漏附近电场太强会导致雪崩击穿, 在沟道长度小于 $2\mu\text{m}$ 的工艺中有时采用缓变扩散 (这将限制最大电场) [Ogura 82]。

沟道长度调制

到此为止, 虽然我们用几种方法计算了 V_{Dsat} , 但所有这些模型在 $v_{DS} > V_{Dsat}$ 时都认为 i_{DS} 为常数。实际的器件情况并不如此, 我们的下一个课题就是研究一种修正来说明这一效应。

从物理上说, 沟道长度的缩短可以说明 V_{Dsat} 以上电流的增加。而这又是合理的, 因为反型层末端与漏之间的耗尽区将随其上电压增加而增加。因此, 虽然反型层末端的饱

和电压不变,但加上电压的距离缩短了。这使得反型层中各处的驱动电场都加大了,因而电流增加。换句话说,我们从电学上加大了晶体管的有效形状因子。这与下述观察事实是一致的:即其它所有条件相同时,短沟器件的微分电导 (di_{DS}/dv_{DS}) 大于长沟器件。

设修正后的沟道长度值为 L' , 则

$$L' = L - A \quad (2.145)$$

其中 A 为耗尽区在漏附近的宽度。根据方程 (2.24) 有

$$A = \sqrt{\frac{2\epsilon_{si}}{qN}(v_{DS} - V'_{DSat})} \quad (2.146)$$

其中 $v_{DS} - V'_{DSat}$ 为耗尽区两端电压。注意,当器件衬底杂质浓度较高时,这一效应减小。

方程 (2.146) 所预计的 I_{DSat} 以上 i_{DS} 的变化是太大了。这主要是由于速度饱和的影响计算不充分。方程 (2.146) 的另一个问题是没有计算耗尽区中的电子,这种电子将使有效的 N 比掺杂浓度高。在例 2.5 中将提出一个较精确的模型。

当问题的解析解很困难时,另一种方案是找出一个能说明特性的经验公式。有一个饱和漏电流的经验公式是

$$i'_D = i_D \left(1 + \frac{v_{DS} - V_{DSat}}{V_A + V_{DSat}} \right) \quad (2.147)$$

其中

$$V_A \approx \epsilon_A L \sqrt{\frac{N_A}{N_T}} \quad (2.148)$$

这里 $\epsilon_A = 5 \text{ V}/\mu\text{m}$, $N_T = 10^{15} \text{ cm}^{-3}$ (Merkel 77)。

经验公式主要有两个局限。首先,它不给出任何物理机理。其次,也很难确定公式的有效范围。不过,实际的模型往往把解析模型和经验的模型结合在一起。

例2.5 沟道长度调制的电路模拟模型。

虽然 2.3 节中的讨论足以获得沟道长度调制的某些感性认识,但我们承认它不是很准确的。在电路模拟中需要一个精确的模型。我们来看看 SPICE2G 是怎样模拟这一现象的。Baum 和 Beneking (Baum 70) 对 A 值作了仔细讨论。解一维泊松方程,他们得到

$$A = \sqrt{\left(\frac{\epsilon_{\text{max}}}{2a} \right)^2 + \frac{v_{DS} - V'_{DSat}}{a}} - \frac{\epsilon_{\text{max}}}{2}$$

其中

$$a \equiv \frac{qN_A}{2\epsilon_{si}}$$

这比方程 (2.146) 精确,但它仍存在两个问题。首先,为计算 V'_{DSat} 必须知道 A , 而为计算 A 必需知道 V'_{DSat} 。这就是说,在电路模拟程序的内环,我们有一对联立的非线性方程。由于这用迭代法来解费时过多,而解析解又是不可能的,我们用 V_{DSat} 值来近似 V'_{DSat} 。那是假定 $A = 0$ 计算出来的。这结果是相当精确的,因为 $|V'_{DSat} - V_{DSat}|$ 通常不超过几百毫

伏。第二个问题是，漏附近的实际几何结构是二维的，而我们在推导模型方程时用的却是一维近似。因此，引进了几个经验因子以保证曲线拟合。在 SPICE 程序中，替代的情况是[Vladimirescu 80]

$$N_A \longrightarrow \text{NEFF} N_A$$

$$V_{DS} - V'_{Dsat} \longrightarrow \text{KAPPA} (V_{DS} - V_{Dsat})$$

由此可见，电路模拟中的器件模型是个相当复杂的问题，因为精确性、计算效率，以及我们获得有关重要现象的分析表达式的有限能力等因素都交错在一起。

短沟效应中的阈值变化

至此为止，我们所有的分析都局限于一维情形，或是在垂直方向或是在横向，这暗含着所有场效应都是正交的这层意思。这一假设在沟道区边缘不成立。当沟道很短或很窄时，有些效应可能成为整个器件特性的重要成份。在本节中我们将考察由于沟道很短而引起的一些效应。

最重要的短沟效应之一是 V_{TE} 随沟道缩短而减小。这主要是由于沟道下耗尽区电荷的电力线横向终止到源和漏区，而不是垂直指向表面。克服这一问题的一个最简单的方法是把二维场问题近似为等效的一维问题。它的办法是把与该横向场联系的电荷从总电荷中减掉，

$$Q'_s = Q_s - Q_L \quad (2.149)$$

这里 Q_L 为等效侧向有关电荷。 Q'_s 是结深及结形状的函数，也是源、漏电压的函数。在评论计算 Q_s 的诸方法时，要记住这一方法只是实际的二维（或三维）问题的一种近似。还应注意，当 L 变大时， $Q_s \gg Q_L$ ，这时 Q'_s 接近于 Q_s 。在 Q_s 与 V_{TE} 之间， T_{ox} 也起着重要作用。

首先采取这种方法的是 Yau[Yau 74]，他把耗尽区的形状近似为一个梯形，得到如下结果：

$$Q'_s = Q_s \left[1 - \left(\sqrt{1 + \frac{2X_{pmax}}{X_j}} - 1 \right) \frac{X_j}{L} \right] \quad (2.150)$$

其中 Q_s 是用传统方法计算的电荷， X_j 为结深， X_{pmax} 为沟道对衬底的正常耗尽区宽度。有关的几何形状示于图 2.28。在 V_{DS} 很大的条件下，梯形会变为三角形，使得上述方程不再成立。注意在方程中，结深 X_j 起着对各种距离归一化的重要作用。这样，一个器件是否短沟器件就不仅取决于实际的沟道长度，更主要的是取决于沟道长度与结深之比。一个用磷扩散作源漏的 $5\mu\text{m}$ 器件所表现的短沟特性可能比以砷作源漏的 $3\mu\text{m}$ 器件

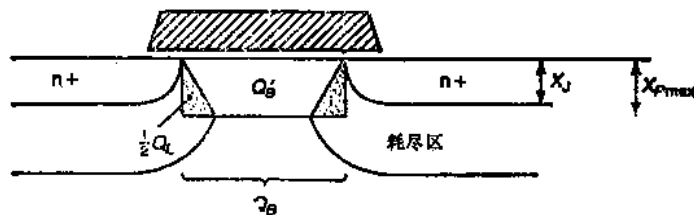


图2.28 短沟效应电荷模型。沟道下耗尽层的部分电荷 Q_L 被认为属于源或漏

更严重。这是因为砷的结深会比磷的结深小得多。

Dang(Dang 79)修订了 Yau 的模型,他假设与源和漏相联系的耗尽区形状为圆柱面。这一关于阈电压的更精确的表达式被用于 SPICE2 中的第三级模型。

注意体效应因子 γ 正比于沟道下的电荷。我们有

$$Q_B = -\gamma C_{ox} \sqrt{v_{SB} + 2\Phi_{FP}} \quad (2.151)$$

由于短沟效应,这一电荷减小,因而有效体效应因子也减小。我们定义 γ' 为

$$\gamma' = \gamma \frac{Q'_B}{Q_B} \quad (2.152)$$

这样,在其它因素相同的情况下,短沟器件的体效应比长沟器件小。

漏感应势垒下降及源漏穿通

在我们分析阈电压的短沟和窄沟效应时假定了源与漏是同电位的。如果假定电位不同,结果会很复杂。漏会作为第二栅极而调制沟道下的电量 Q_n 。这使得阈电压明显地随漏电压漂移,因为漏要调制沟道中的势垒。我们引入一个经验参数 σ ,它直接作用于 V_T ,如下式所示:

$$V'_{Tn} = V_{Tn} - \sigma v_{DS} \quad (2.153)$$

其中 σ 为静电场反馈的系数。Ratnakumar, Meindl 和 Scharfetter(Ratnakumar 81)提出

$$\sigma = \frac{6T_{ox}}{X_{Fmax}} \exp\left(\frac{-\pi L}{4X_{Fmax}}\right) \quad (2.154)$$

在离子注入晶体管的情况中,我们把 X_{Fmax} 解释为沟道下的耗尽区宽度¹⁾。

漏感应势垒下降对工作在阈值附近的短沟器件是影响很大的效应。这样的器件工作在饱和区时, σ 是决定输出电导的主要因素。在第四章中我们将看到漏感应势垒下降具有重要影响的两个电路例子。漏感应势垒下降限制了 MOS 反相器的最大增益。它对决定有比和预充电电路的噪声容限也有重要作用。

当 v_{DS} 足够大时可能发生穿通。在发生穿通时,通常在次表面²⁾有一个基本不受栅控制的电流在源漏之间流过。由于同样的机理,栅电压使硅表面反型,而漏电压可以使沟道下面的势垒降低。当势垒降得足够低时,电流可以从这一通道流过。因为这一问题本质上是二维的,所以很难得到穿通电压 V_{PT} 的准确方程。

粗略地说,穿通电压随杂质浓度线性增加而随 L 以平方关系增加。当 V_{DS} 的幅度增加时,次表面穿通电压也增加³⁾,这是因为势垒增加了。另一种增加次表面穿通电压的办法是进行一次沟道下的深注入。但这有副作用,即,使得体效应增加。

次开启导电

在器件方程的推导中,假定了晶体管的导通是突然发生的,这时源对沟道的势垒减小了 $2\Phi_{FP}$,或者说沟道电位减小了 $v_{SB} + 2\Phi_{FP}$ 。这一条件被用来定义 V_{Tn} 。在实际器件

1) SPICE2G, level 3 中,取 $\sigma \propto L^{-3}$ 。

2) 只要电位分布满足,穿通可在任何深度发生。

3) 当耗尽区相遇时即发生穿通,这个通俗说法是错误的。

中, 导通并不完全是突然的, 在 V_{TS} 以下, 器件的 i_{DS} 与 v_{DS} 有一指数关系。

当沟道电位在 $v_{SB} + \phi_{Fp}$ 与 $v_{SB} + 2\phi_{Fp}$ 之间时, 晶体管处于弱反型状态。有两种现象影响源漏电流。第一个是, 沟道中有足够能量导电的电子数粗略地与 $\exp(q\phi_s/kT)$ 成比例。这些电子能够越过源到沟道的势垒而参与 i_{DS} 。所以, 这一工作区的 $I-V$ 指数特性是与正向偏置 p-n 结的 $I-V$ 指数特性相仿的。

第二个重要考虑是 v_{GS} 对 ϕ_s 的杠杆作用。 v_{GS} 增加趋向于阈电压时, v_{GS} 对 ϕ_s 的控制能力急剧下降。所以我们可以期望 i_{DS} 随 v_{GS} 的变化比随 ϕ_s 的变化缓慢。考虑某一级效应, 我们可以预期源电流与 $\exp(qv_{GS}/nkT)$ 成比例, 其中的 n 反映这样一个事实: v_{GS} 改变 10mV 不足以使 ϕ_s 改变 10mV。

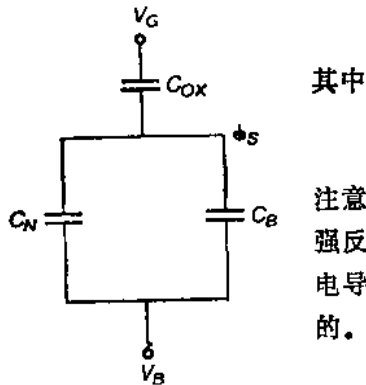
通常, 较仔细地计算电荷可以改进模型。按照 Swanson 和 Meindl [Swanson 72] 的工作, 我们提出一个这一工作区的小信号模型。设 C_N 为反型层电子的电容, C_B 为体耗尽层的电容,

$$C_N = \frac{\partial Q_N}{\partial \phi_s} \quad (2.155)$$

$$C_B = \frac{\partial Q_B}{\partial \phi_s} \quad (2.156)$$

采用这些电容的一个电路模型示于图 2.29。在弱反型区, C_N 比 C_B 小得多。根据这一模型, 我们发展了阈电压概念, 将它定义为次开启导电和强反型之间的分界, 有

$$V'_{TS} = V_{TS} + \frac{nkT}{q} \quad (2.157)$$



其中

$$n = \frac{C_B + C_{Ox}}{C_{Ox}} \quad (2.158)$$

注意这一阈值并不是实际意义上的电路设计阈值。它主要是强反型与弱反型的一个交接点。在 V'_{TS} 下, 可能有相当大的电导。虽然 n 有一定物理意义, 但它通常是由实验数据决定的。

我们已经指出, $v_{GS} < V'_{TS}$ 时, 电流可表示为

$$i_{DS} = I_0 \exp \left[\frac{q(v_{GS} - V'_{TS})}{nkT} \right] \quad (2.159)$$

图 2.29 弱反型晶体管的电容模型

有一个对设计很有用的品质因子是 $dv_{GS}/d \log I_{DS}$, 它表示将漏电流减小一个数量级需要减小的栅电压。我们有

$$\log I_{DS} = \log I_0 + \frac{q \log e}{nkT} \left(V_{GS} - V_{TS} - \frac{nkT}{q} \right) \quad (2.160)$$

微分可得

$$d(\log I_{DS}) = \frac{q \log e}{nkT} dV_{GS} \quad (2.161)$$

定义 α 为次开启优值, 有

$$\alpha \equiv \frac{dV_{GS}}{d(\log I_{DS})} = \frac{nkT}{q \log e} = 2.3 \frac{nkT}{q} \quad (2.162)$$

典型的 n 值在 1.0 到 2.5 的范围。由于 2.5°C 时 kT/q 值为 0.0257V ，室温下 α 值通常在 60 到 $150\text{mV}/10$ 倍， 100°C 时 α 值为 $75\text{—}185\text{mV}/10$ 倍范围。要把晶体管完全关断可能需要几百毫伏，这时晶体管的源-漏漏电流与源漏与衬底的 $p\text{—}n$ 结反向偏置漏电流具有同一量级。

短沟效应使次开启电流加大。Brews 等人 [Brews 80] 通过大量实验和二维模拟发现，具有长沟次开启特性的最短器件沟道长度（以 μm 表示）可由下面经验公式表示：

$$L_{\min} = 0.4 [X_j T_{\text{ox}} (X_s + X_D)^2]^{1/2} \quad (2.163)$$

其中 X_j 为结深（以 μm 表示）， T_{ox} 以 $\text{\AA}$ 表示， X_s 、 X_D 分别为源和漏的耗尽区深度（以 μm 表示）。当 L 小于 $L_{\min}$ 时，次开启电流将增加，这是由于源、漏耗尽区的侵占，有效的沟道长度减小了。图 2.30 示出了两个不同 nMOS 器件的次开启特性。衬底杂质浓度及沟道长度的影响是明显的。

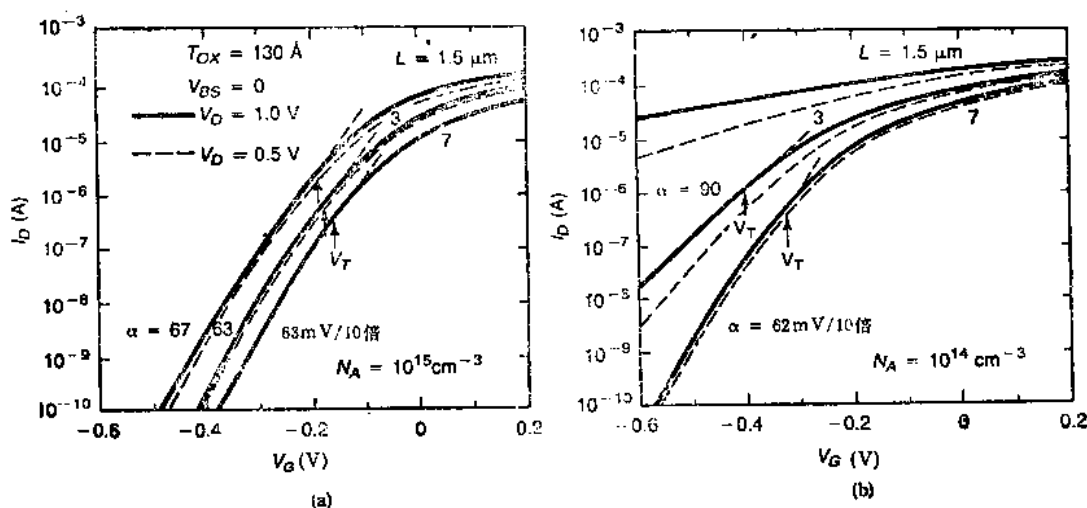


图 2.30 不同沟道长度的晶体管的次开启漏电特性
(a) $N_A = 10^{15}\text{cm}^{-3}$, (b) $N_A = 10^{14}\text{cm}^{-3}$ [Sze 81, p. 470]

离子注入沟道

离子注入提供了一种控制器件阈值而几乎不改变其它器件参数的方法。阈电压和体效应参数都与衬底杂质浓度有关，当然还有其它因素。离子注入技术使得人们可以有选择地控制 N_A 。人们可以在芯片的不同区域改变 N_A 值并精确控制掺杂分布。由于离子注入是一种低温工艺，许多材料，包括光致抗蚀剂，都可以用作掩蔽注入杂质。

加速电压及注入杂质的特性决定了杂质的注入深度。沿深度的分布可以用高斯分布描述，但如果需要，可以通过随后的扩散修正这一分布。分布的特征可以用高斯分布峰值的深度 R_p 及分布的标准偏差 ΔR_p （也称为“弥散”）来表述。我们把注进硅片表面每平方厘米的总原子数定义为注入剂量。硅和二氧化硅中不同原子在不同能量下的 R_p 和 ΔR_p 测量值已被列出表来 [Gibbons 75]。我们定义 R_p 处的杂质浓度，也就是杂质浓度峰值为 N_p ，而背景浓度为 N_b 。将高斯分布下的杂质分布积分，可得

$$N = \sqrt{2\pi} \Delta R_p (N_p - N_b) \quad (2.164)$$

注入后还需要高温退火。这使得晶格松动，由离子束造成的损伤可以消除，并且使

杂质可以就位到晶格位置。

在 VLSI 中注入晶体管是很普通的。对 nMOS 增强型器件常用受主杂质轻注入，而对 nMOS 耗尽型器件则常用较重的施主杂质注入。CMOS 中两种器件通常都要采用注入。阱中器件的注入常常是用来降低有效的杂质浓度而不是增加杂质浓度。例如，阱注入可以被用来部分抵消阱中的重硼注入。

我们首先看看增强型 nMOS 器件的情形。具体说，我们看看用在硅表面进行轻硼注入来增加器件阈电压的情形。注入的结果是，一个高斯分布的 P 型杂质区叠加在均匀轻掺杂的衬底浓度之上，如图 2.31 所示。正确计算电荷比准确模拟其分布重要得多。这是因为电荷影响端电压是通过泊松方程中的积分来实现的。这一积分就将分布形状的细节滤掉了¹⁾。因此，我们用一个方形来近似高斯分布。这看起来很粗糙，但要注意高斯分布的尾巴下降得很陡。图 2.31 也画出了理想化了的注入器件的分布形状。现在问题简化为沟道下只有两级不同掺杂水平的情形。

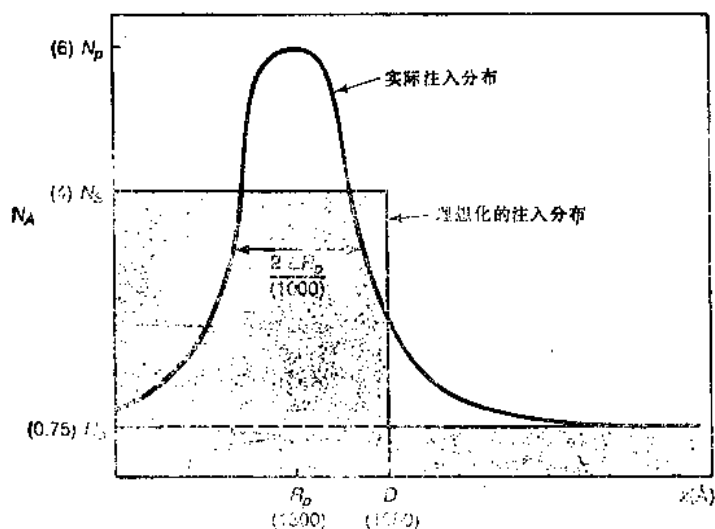


图 2.31 注入的增强型晶体管 (括号中的数字取自例题)

要强调的是，注入对平带电压没有影响，可以从很多方面说明这一点。通过对化学势应用基尔霍夫电压定律，我们知道如果在硅表面与 P 型体内深处之间放置任何导电材料，而体内被接地，则这一材料的影响将被抵消掉， V_{FB} 不会改变。也可以从另一途径看这一问题，设硅表面的掺杂浓度比我们原来的参考点高。在这种情况下我们可能得出结论认为我们原来计算的平带电压不对，因为 V_{FB} 与 ϕ_F 有关。但这只有在我们以较重掺杂的硅表面为参考来测量平带电压时才是对的。另一方面，这一电压误差正是较重掺杂的表面与较轻掺杂的衬底之间的电位差，而衬底电位是我们取的参考点。这样一切都被抵消掉了。

当晶体管的栅压从 V_{FB} 向 V_T 增加时，硅-二氧化硅界面下的区域将从表面开始变为耗尽。为了符号使用的方便，我们将等效的方块表面掺杂记为 N_s ，相应的体效应因子

1) 另一方面，微分参数确实与注入分布形状的细节有关。

为 γ_s ，方块向体内延伸的距离为 D 。而在体内深处，参数为 N_A 和 γ_A 。一个关键的问题是 D 是否比耗尽区宽度 X_{pmax} 大。我们可以计算出 $X_{pmax} = D$ 时的源电压 V_{SB} ：

$$V_{SB} = \frac{qN_s D^2}{2\epsilon_{si}} - 2\Phi_{Fps} \quad (2.165)$$

其中 Φ_{Fps} 为对表面掺杂 N_s 的 Φ_{Fp}^D 。若 V_{SB} 为正，则 V_{Tx} 仅仅与 N_s 有关。在这种情况下有

$$V_{Tx} = V_{FB} + 2\Phi_{Fps} + \gamma_s \sqrt{2\Phi_{Fps}} \quad (2.166)$$

这样我们得到了一个很直观的结果：若所有起作用的电荷都在方块内，则器件完全像整个衬底都掺杂到 N_s 的样子。不过，即使是这种情形注入器件也有两个重要的好处。首先，只在沟道下杂质浓度是高的，源、漏连线下面的结电容比整个衬底重掺杂的情形低得多。此外，对于源跟随器结构²⁾，当源电压超过方程 (2.165) 给出的临界值时，体效应将小很多。对于这种情形这是很需要的，因为这可以使源跟随器的电压增益接近于 1。体效应参数与源电压的关系示于图 2.32。在大多数离子注入器件中，如前所假设， D 是大于 X_{pmax} 的。如果情况不是这样，则必须把问题分成两部分。在耗尽区达到 D 以前，我们的问题和上面讨论的一样。但超过该点后，我们必须采用 N_A 而不是 N_s 。

在 nMOS 工艺中，可以采用非零的偏压 V_{SB} 使耗尽区边缘处于 γ 值低的情况。也可能有不同的情况，例如，在典型的 $2\mu m$ p 阱 CMOS 工艺中，若 p 沟器件的源偏离 V_{DD} ，则其体效应降低，而若阱中 n 沟器件的源电压增加，则其体效应上升。

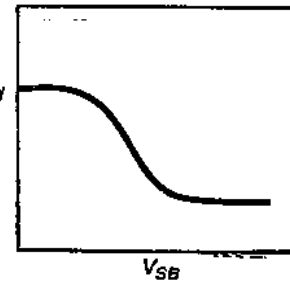


图 2.32 一个离子注入器件的体效应与源电压的关系

不幸的是，大多数电路模拟程序，包括 SPICE2，不允许 γ 值为非常数。为了适应这种不足之处，我们必须利用可供调整的参数得到与实际器件曲线合理的拟合。如果 D 近似等于 X_{pmax} ，则在广阔的工作区， γ 取相应于 N_A 的 γ_A 值。但若我们采用这一值而不作任何其它调整，则模拟程序将计算出错误的 V_{Tx} 值，而这正是电路设计中的一个关键参数。考虑这一点，我们可以调整 V_{FB} 值以获得适当的 V_{Tx} 结果。综合调整这些参数可以得到相当不错的与电流特性吻合的结果。注意，在电路设计中，虽然 V_{FB} 不是关键参数，它却对 MOS 电容模型有影响。当模拟的电路在 V_{FB} 附近对电容变化有较高的灵敏度时，必须考虑这一点。

离子注入的一个很重要的用途是制作耗尽型 MOS 器件。在 n 沟工艺中，或是砷或是磷被注入到沟道区以形成耗尽型器件。

为了能够分析耗尽型器件，需要做比增强型情形更大的简化。为了计算 V_{Tx} ，我们需要求出刚好 $Q_N = 0$ 时的栅电压，沟道刚刚耗尽*前，沟道区将包含体内耗尽的硼负离子及表面的一层砷或磷正离子。假定正离子是紧靠表面的，它们完全类似于 2.1 节中分析的 Q_f 。与增强型情形不同，这里我们将要求平带电压的平移。实际上，两种情形的能带都不可能完全平，用“平带”一词似乎有点“牛头不对马嘴”。但从线路的观点看，

1) 我们用 $2\Phi_{Fps}$ 还是 $\Phi_{Fps} + \Phi_{Fp}$ 来定义反型的开始并不重要。反型时的表面浓度并不是绝对的。

2) 在这种结构中，源电压是不固定的。

* 原文为“反型”，概念有误，作了更正。——译者注

平带电压是栅对体电容为 WLC_{ox} 时的最大电压。在 P 型衬底上加不同的 P 型材料杂质将不影响这一问题，但加 n 型杂质就要影响了。其理由是，如果我们采用前面关于平带的定义，则当表面没有多余电荷时，表面是 n 型的，不但存在着一个耗尽区改变了电容，而且晶体管也是导通的。所以在简单的分析中，最好是假定我们感兴趣的平带电压确实改变了。改变量由下式给出：

$$\Delta V_{FB} = -\frac{Q_s}{C_{ox}} \quad (2.167)$$

其中 Q_s 为注入的表面电荷密度 qN_s 。

这一分析对浅注入是相当精确的。深注入则需要复杂得多的分析，包括考虑迁移率随与表面的距离而变化。SPICE 不具有这种特点，所以我们必须调整其中某些参数来拟合特性。通常这除了要调整 V_{FB} 外还要调整 γ 。注意与增强型注入不同，现在实际上有 V_{FB} 的平移，尽管它的特性比我们所分析的还要复杂一些。

具有深注入的耗尽型晶体管的关断是困难的。当栅电压向阈值下降时，在硅表面的顶部形成耗尽区。这样，耗尽型器件的沟道实际上是在表面以下，夹在顶部和底部的耗尽区之间，如图 2.33 所示。现在有效的栅介质包括原来的氧化层 T_{ox} 加上顶部的耗尽层¹⁾。这降低了栅电极对 Q_N 控制的效率。除非衬底负偏置把沟道推向表面，耗尽型器件可能严重漏电。



图 2.33 栅反偏条件下的耗尽型器件。表面下存在着沟道

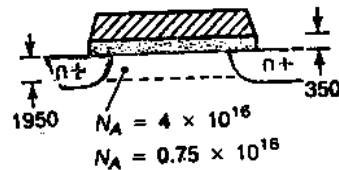


图 2.34 注入的增强型 MOS 晶体管

例 2.6 注入晶体管的阈值计算。

这一注入晶体管例子取自 Dennard 等人 [Dennard 74] 关于按比例缩小的经典论文。这一例中，用 $T_{ox}=350 \text{ \AA}$ ，衬底为 $2\Omega \cdot \text{cm}$ 的 P 型材料。从图 2.19 可以查出这一电阻率相当于 $N_A=7.5 \times 10^{15} \text{ cm}^{-3}$ 。用 40keV 下的硼离子对器件进行注入，剂量为 $6.7 \times 10^{11} \text{ cm}^{-2}$ 。从射程表求出 $R_p=1300 \text{ \AA}$ ， $\Delta R_p=500 \text{ \AA}$ ，估计有 3% 的注入离子损失在氧化层中。我们作粗略的近似，设等效的方块形分布深度 D 为 $1950 \text{ \AA}$ ，并计算出等效的浓度 N_s 为 $4 \times 10^{16} \text{ cm}^{-3}$ 。采用这一杂质浓度，我们计算出反型时 $\phi_s=0.768 \text{ V}$ 。耗尽层宽度 X_d 等于 D 时的外加电压为 $V_s=0.155 \text{ V}$ 。注意在零偏压下，耗尽区完全被包含在注入杂质之内。假设 $V_{FB}=-1 \text{ V}$ ，我们可以用标准的 V_{TX} 方程计算出零偏压下的阈值。由于对这一计算重要的 Q_s 都在注入区内，我们采用对应于 N_s 的 γ 。在注入区内 $\gamma=1.167 \text{ V}^{1/2}$ ，我们有 $V_{TX}=0.79 \text{ V}$ 。我们可以与没有注入的情形作一对比，在那种情形下，反型时 ϕ_s 为 0.681 V ， $\gamma=0.505 \text{ V}^{1/2}$ ， $V_{TX}=0.097 \text{ V}$ 。有关的几何尺寸示于图 2.31 和图 2.34。

现在我们考虑耗尽型晶体管的情形，其中用硼代替硼作为注入杂质。用 1.2×10^{18} 个原子/ cm^3 的剂量得出阈电压为 -1.9 V 。记住要用 N_A ，并且假设体效应不受注入的影响。

1) 当氧化层很薄时，由于反型层有一定厚度，即使对于非注入的器件，有效的 T_{ox} 也比物理上的 T_{ox} 大。

窄沟效应引起的阈电压变化

前面我们引进了一个修正因子 L/W 来帮助说明鸟嘴和场注对有源区侵入的效果。这一修正因子对于精确模拟非常窄的器件的工作是不够的(所谓“窄”与工艺的垂直尺度有关,但粗略地说,对 $2\mu\text{m}$ 工艺, $W < 4\mu\text{m}$ 可以看作是窄的)。

最显著的窄沟效应是窄沟器件阈电压向正的方向改变。这一效应在衬底浓度低而场注浓度高时尤其突出。这种情况下,阈电压的改变主要是由于场注杂质向沟道区扩散使有效的背景浓度增加。由于大多数工艺都属这种情况,我们将集中讨论这一机理。

其它影响窄沟阈电压的机理有由于鸟嘴而引起的氧化物斜坡(它也影响器件的增益)以及在器件边缘垂直场的发散,垂直场发散效应的作用也是使阈电压增加。

当场注杂质的延伸完全跨越了有源区时,就出现侧向扩散的极限情况。这时我们必须把场注杂质、器件阈值调整注入以及衬底的浓度加在一起来计算最坏情况下场注杂质所造成的阈值改变。关于这一问题,就我们所知,现在还没有解析解能够把实际器件中杂质浓度和氧化层厚度的非均匀性都计算在内。

Yang和Chatterjee[Yang 82]提出了一个关于 γ 和 V_{rx} 与 W 关系的经验模型。他们提出

$$\gamma' = \gamma + \frac{\gamma_w}{W} \quad (2.168)$$

和

$$V'_{rx} = V_{rx} + \frac{V_{rxw}}{\sqrt{W}} \quad (2.169)$$

方程(2.169)是基于这样一个模型:它假定 Q_s 从栅氧化层下的理想值到场区的值是线性过渡的。注意, W 减小时, γ 和 V_r 增加;而 L 减小时, γ 和 V_r 都减小。

例2.7 窄沟效应。

考虑例2.6中的增强型注入器件。典型的场注入浓度在 $3 \times 10^{18} \text{cm}^{-3}$ 的范围,这样,对一个增强型窄沟器件,有效的杂质浓度包括三个组成部分,即背景浓度 7.5×10^{15} ,器件注入浓度 4×10^{18} ,以及场注入 3×10^{18} 。当场注入杂质完全扩散到晶体管里时, γ 则从 $1.167 \text{V}^{1/2}$ 增加到 $1.49 \text{V}^{1/2}$ 。这又使得阈值增加 238mV ,达到 1.073V 。注意场注杂质在场氧化的长时间过程中也扩散进衬底相当深。这样,即使加上源偏置,这一较高的 γ 值也倾向于保持不变。

对于耗尽型器件, γ 将从 0.505 增加到 1.13 , V_{rx} 从 -1.9V 增加 550mV 达到 -1.35V 。增加的量比增强型大,这是由于有效的起始浓度较低。

体CMOS工艺中的自锁效应

在体CMOS工艺中有一个固有的从 V_{DD} 到地的四层 $p^+/n/p/n^+$ 通路。这一pnpn通路包括三个 $p-n$ 结,它们构成了一对交叉耦合的pnp和npn双极晶体管,对n阱工艺,如图2.35所示。这两个晶体管形成了一个正反馈通路,每个晶体管的集电极驱动另一晶体管的基极,如图2.36所示。

本节中我们将采用n阱工艺进行讨论。只要把电源的极性以及n型和p型反过来,就可以得到相应的p阱工艺情形。

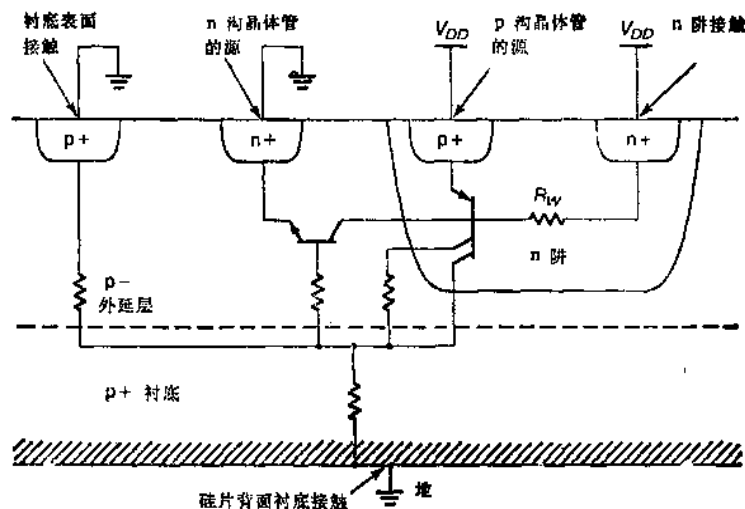


图2.35 体CMOS中固有的自锁通道（为清楚起见省略了n沟及p沟晶体管的栅及漏，而只画出源。有的工艺设计中可能不用衬底接触，而利用封装方法中芯片背面接触）

pnp晶体管 M_w 是一个垂直晶体管，其电流放大系数 β 可在50到几百的范围。 M_s 为横向npn晶体管，以衬底为基极，阱为集电极。这一晶体管的 β 通常在0.5到10的范围，它随着 n^+ 发射极与阱的间距增加而指数下降为零。

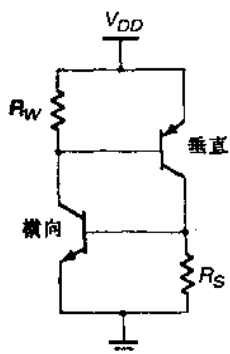


图2.36 相应于图2.35 反馈通路的自锁回路

电阻 R_w 和 R_s （如图2.36所示）分别把它们相应的晶体管的基极和发射极旁路， R_w 一般在1k Ω 到20k Ω 的范围。衬底电阻 R_s 与衬底情况有很大关系：可能整个衬底为 p^- 材料；也可能是生长在重掺杂 p^+ 衬底上的 p^- 外延层。在 p^-/p^+ 外延的情形，衬底起接地平面的作用，从而使 R_s 的值降低到几个欧姆，而通常 R_s 可以高达500—700 Ω 。

如果两个晶体管的电流放大系数以及两个电阻 R_w 和 R_s 的值足够高，则电路可能被外界的扰动触发进正反馈的状态，在这种状态中每个晶体管不断地驱动着另一个晶体管。两个晶体管的电流都将不断加大直至受到自身限制或使芯片烧毁或使压焊线烧断。这种情况就称之为自锁。

即使发生自锁效应的pnpn通道没有使芯片烧毁，在大多数情况下也会使逻辑电路功能混乱。为了把一个芯片恢复到非自锁的状态，通常要去掉电源（使 V_{DD} 为零），然后重新通电。

对某些工艺设计或版图结构，单个的自锁pnpn通道本身可能不是破坏性的，这是因为其电流被限制在安全的水平内。不过，一对自锁的晶体管的扰动可能会使本来“安全”的其它邻近pnpn通道发生自锁。这样，由于自锁可能传播到芯片的大部分区域，本来非破坏性的情况会导致失效。

因为自锁的后果是这样地严重（导致失效或逻辑混乱），为了采取能够完全消除自锁可能性的技术措施就需要了解什么样的电路干扰可能引起自锁。下面我们列出一些可

能触发自锁的常见干扰及机制。

通过信号输入/输出引脚的干扰

传输线反射或阻尼振荡：若在输出母线或传输线上的反射或阻尼振荡使得一个 p 沟晶体管的 p^+ 源或漏的电压比 V_{DD} 高出约 0.6V，则以 p^+ 区为其发射极的垂直 pnp 晶体管将导通。若这一晶体管本身是一个有可能自锁的环路的组成部分，就可能导致自锁。当一个 n 沟晶体管的 n^+ 源或漏的电压低于地时，也将发生类似情况。

带电接插电路板或模块：若一个未加电源的电路板被插进一个已经通电的系统，（例如要更换一个有毛病的模块），由于新插入的电路板的 V_{DD} 要较慢地上升到稳定值，在芯片的信号端可能看到比其 V_{DD} 高出 0.6V 的电压。当芯片充分通电时，除非采取了适当的抑制措施，芯片的一部分将被自锁，当然，若不允许带电接插模块也就不会发生这种情况。

静电放电 (ESD)：静电放电打火（往往来源于干燥环境下的人体）能够触发自锁。这一电荷脉冲通常将进入输入压焊块，而这一压焊块是由 ESD 保护电路箝位的。箝位电路的器件将把少数注入到衬底或阱中，有可能触发邻近的其它 pnpn 通道。若 ESD 保护电路防护得不够好，则脉冲过后电路将被自锁。ESD 保护电路的设计将在 5.7 节中讨论。

在芯片内部或者通过 V_{DD} 或地产生的干扰

在接通电源时 V_{DD} 的摆动由于阱电容的存在可能在阱和衬底产生容性位移电流。在阱中这一位移电流的影响比在衬底中更坏。这是因为所有的位移电荷都必须通过少量的阱接触（通到 V_{DD} 的欧姆接触）流出，而对衬底来说，电荷对地通路的电阻相对说较低。阱中的侧向压降可能使阱中垂直双极晶体管之一导通，从而导致自锁。

由于大量内部驱动器的开关，在电源线或地线上产生突然的瞬态电压，可能使某个 pnpn 通道的双极晶体管导通，导致自锁。在具体情况下这能否发生，应加以仔细检查。

阱结的泄漏电流可能产生足够大的侧向压降使阱中的垂直双极晶体管导通。

辐射如 X 射线、宇宙射线，或 α 射线等可能在粒子经过的阱区及衬底区产生电子-空穴对，这有可能触发潜在的 pnpn 自锁通道。

在衬底 n 沟器件附近由热电子感应的少数电流被附近的阱收集，从而使这个阱中的垂直晶体管导通，由此触发一条 pnpn 通道。此外，多子衬底电流也会使衬底电位增加以致使附近的 n^+ （源或漏）有源区导通。下面我们将讨论用一个 p^+ 衬底上的外延层将完全消除这一多子效应。

这里将不推导自锁的精确的必要充分条件。文献中对此有几种不同的处理方法 [Estreich 82, Gregory 73, Schroeder 80]。自锁的一个必要条件如下：

要把一个 pnp/npn 晶体管对触发成自锁状态，触发的干扰必须把器件偏置到反馈回路微分增益大于 1 的状态。

这一条件对我们研究防止自锁的措施是很有用的。

防止自锁的措施，不论是靠工艺设计还是靠适当的版图设计或屏蔽技术，其目的通常都是防止在给定干扰下环路微分增益大于 1。这有两种实现方法：或者是靠阻断或衰

减干扰,或者是靠衰减回路微分增益使得晶体管即使被干扰导通了,环路微分增益也不会大于1。

由于自锁有如此严重的后果,就必须在工艺结构设计或在电路版图横向尺寸设计方面采取措施,使它在给定的系统环境下永远不会发生。下面我们讨论四种通常的抑制措施。

降低 β : npn/pnp晶体管对要自锁,两个(双极)晶体管的 β 的乘积就必须大于1。这一条件是由上述必要条件导出的,并且还是一个比较弱的条件。但由此确实可以想到一种防止自锁的策略:采用两个 β 乘积不会大于1的器件结构(靠适当的纵向工艺设计及横向间距)。这一策略的问题在于,VLSI中器件的纵向结构和侧向间距往往使 β 乘积远大于1,要使侧向间距加大到足以使 β 乘积充分下降,往往使电路密度就低到不能满足要求了。因而降低 β 对大多数工艺来说最多也只是部分对策。

电阻分流:最有效的对策(有时与降低 β 相结合)是使电阻 R_s 和 R_w 足够小以保证把可能出现的基极电流旁路到电源干线中去。通过使这两个电阻或其中之一足够小,至少可以使一个晶体管的基极对发射极电压保持在开始导通的0.6V以下。只要防止一个晶体管导通,回路增益就不可能超过1,这就抑制了自锁。 R_w 是靠足够好的阱接触来减小的,这种阱与 V_{DD} 的欧姆接触通常是通过金属完成的。为了达到很低的 R_w 值,要在面积上作些牺牲,可以在n阱周围加一个n⁺环(对n阱工艺)。这可以大大减小 R_w (从几k Ω 减到小于100 Ω),这在I/O电路中特别有用,因为这种电路中干扰最严重。

在非外延工艺中,为了减小 R_s ,可以采用频繁接引出线(对地的欧姆接触)的方法。这些接触的适当空间频度通常由设计规则给出。阱与衬底接地点的空间位置的确定是个比较困难的CAD工作。

采用重掺杂衬底上的外延层可以大大减小 R_s ,这被认为是防止自锁最有效的办法。虽然单独使用外延层(也就是说没有其它措施)还不能完全防止自锁,但把它与其它措施结合起来就使得自锁的控制比较容易了。

采用外延的办法来控制自锁,至少对低至1.5 μm 的工艺来说是满意的。线宽更小(n⁺与p⁺的间隔也成比例降低)时,则需要采取其它预防措施。

如果使重掺杂衬底上外延的硅片背面接地,则很少需要在外延层表面做接触了。不过在某些封装方案中排除了背面接触的做法,在这种情况下就需要表面接触。即使是这样,它也比非外延工艺好得多。表面的环形接触几乎与背面接触一样有效。

有些方案靠薄层电阻较低的特殊阱结构也可以降低阱电阻 R_w 。一种办法是用深注入来达到反转阱的杂质分布。这使阱的薄层电阻降低,与此同时,为了表面上的器件的需要,维持着适当的表面浓度。文献中也描述了采用埋层的办法来达到同一目标[Estreich 80, Walczyk 83],在与双极兼容的CMOS工艺中常考虑埋层方案。

其它工艺结构:对于亚微米工艺,需要采取更重大的步骤。这些步骤可能有:沟槽隔离[Rung 82],它用一个足够深可抵达低阻衬底的氧化沟槽把阱与衬底区隔离开来;绝缘体上的硅(SOI),它完全消除了pnpn通道,因为它的体区域或是在绝缘体上面或是互相之间被绝缘体隔离开(而不是反向偏置的pn结);还有其它的介电隔离方案,它们也都消除了pnpn通道。

在电路及版图设计时采取的预防措施:上面抑制自锁的措施都是在基本工艺结构上

面进行适当设计。但在进行电路和版图设计时也应采取附加的措施。现在就讨论这样一些措施。

要求最严格的横向间距是 n^+ 和 p^+ 有源区之间的距离(源或漏对接地欧姆接触)。由于阱结也处在 p^+ 和 n^+ 有源区之间,降低其间距对于达到VLSI所需要的高电路密度是重要的。在双阱工艺中,二阱的边缘都处在 p^+ 和 n^+ 有源区之间。 n^+ 和 p^+ 的间距太小会导致阱的穿通。这会引起高的漏电流,从而会引起自锁。侧向晶体管的 β 也会由于间距减小而急剧增加。这些因素非常复杂而且与工艺有关,这里不准备作进一步的详细讨论,但要说明这是设计考虑中的关键因素。粗略地说, n^+ 与 p^+ 的间距可以小到与外延层厚度同一数量级。

可以在版图的“基础层”采取强有力的自锁抑制措施。如上所述,最严重的自锁干扰是从信号I/O压焊块进入的。因此可以在这些I/O区周围加上强保护。这种保护包括在大多数阱周围加阱环(前面已讨论过)以及在可能将少数载流子发射到衬底及外延层的区域周围加收集“阱”(由一条阱带组成)。

若工艺不采用外延层,则可以将衬底的接地条带(与接地的金属相接触的 p^+ 扩散区)加在阱结构的外围,使由阱流出来的多子短路到地。

如果在整个芯片上都采用这种保守的I/O自锁抑制措施,就不可能达到VLSI所需的高电路密度。因此要使得内部区域的 p^+ 与 n^+ 有源区间距减小,而其保护放松,由此达到高的电路密度。

在这种情形中,存在着许多pnpn通道,如果允许干扰抵达这些通道,就会导致自锁。自锁是靠严密防护这一内部区的边界,使得不可能有干扰通过衬底进入这一内部区来防止的。其原理图示于图2.37。在外围的器件被严密防护,以衰减可能通过I/O信号压焊块进来的比较严重的干扰。内部区由一个环形保护区屏蔽起来,如图2.38所示。这样,唯一可能进来的干扰是那些在 V_{DD} 及地线上的干扰,以及那些以辐射形式进来的干扰。

需要注意的外围电路包括:ESD“二极管”(当信号冲过 V_{DD} 或地时,它可能向衬底或阱区发射少子);输出器件(特别是那些处于关断状态的三态器件,在这种器件中输出漏区是悬浮的);以及传输门(这种传输门也可能有悬浮漏区直接连向I/O压焊块)。

集电极保护结构工作情况如下。从 n^+ 发射区发射的少子电子被 n 阱集电区截取。由衬底与外延层间 p^+-p^- 梯度产生的自建电场对少子电子施加作用力使少子保持在外延区,这些少子电子可以被保护阱收集。对这种可能导致自锁的电子的衰减是非常强的,约为一百万比一[Troutman 83]。

围绕内部区的完美保护结构可能完全阻止干扰从衬底进来。不过有些干扰可能从其它途径进来,包括 V_{DD} 的干扰及辐射效应。

显然, V_{DD} 电源干线必须进到内部区域以便向内部电路供电。接通电源, V_{DD} 从零向满度值上摆时,与 dV_{DD}/dt 成比例的容性位移电流将从阱及衬底流过。通常,阱区位移电流的效应较强。

在这种情形中,当阱中的位移电流大到其压降能够使一个阱垂直晶体管导通时就会触发自锁,这个垂直晶体管的导通导致了阱外一个侧向晶体管导通,从而引起再生和自

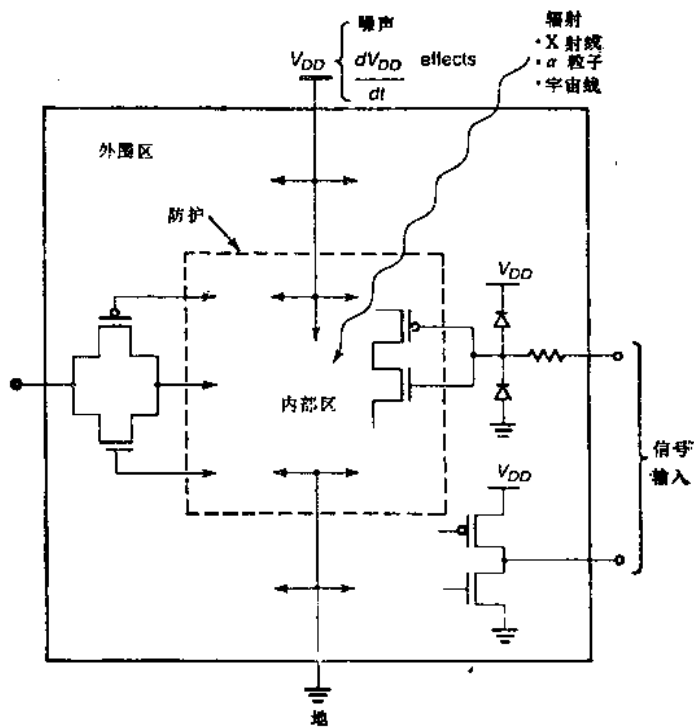


图2.37 分割内部区与外围区的原理图

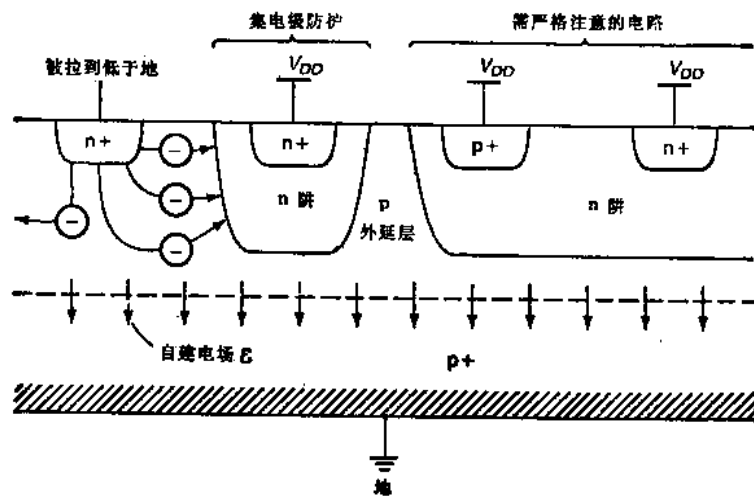


图2.38 集电极保护结构

锁。这种方式的自锁可以通过在阱中增加大量接触（对n阱来说，连到 V_{DD} ）来防止。从任何一点到 V_{DD} 的电阻要足够少使其压降不会使基极-发射极结导通。

外界辐射，如X射线、 α 粒子、宇宙线等，可能穿透芯片并沿其轨迹产生许多电子空穴对。在某些情况下，这种干扰可能相当严重以致引起自锁。这种辐射曝光的一个例子是，一个计算器中的CMOS芯片在通过机场安全检查时要接受X射线辐射。对这种情况可以采取大量阱接触及适当的衬底结构（或是用大量衬底接触，或是使用外延层）来

防止自锁。由于 α 粒子轰击引发的电流尖峰时间是非常短的，通常认为，从理论上说即使其幅度大到可能引起自锁，但由于脉冲非常窄，也不会使回路不稳定。

二级效应小结

ΔW	宽度侵蚀。 W 减小使 I_D 减小， V_T 和 γ 增加。
ΔL_{poly}	多晶硅栅长的变化。 L 缩短使 V_T 、 γ 减小， I_D 、 σ 增加。
X_j	结深。门、漏结深较大时，短沟效应较严重。
L_{diff}	侧向扩散。 L_{diff} 增加时 L 减小， C_{gs0} 和 C_{gd0} 增加。
C_{gs0} , C_{gd0}	栅重叠电容。
μ_0	本征迁移率。 I_D 与 μ_0 成比例变化，温度 T 增高时 μ_0 减小。
f_v	垂直场迁移率缩小因子。 V_{GS} 增大时 μ 及 I_D 减小。
f_h	水平场迁移率缩小因子。 V_{DS} 增大时 μ 减小。
v_{max}	载流子饱和速度。使 I_D 在较低的电流下达到饱和， I_{Dsat} 偏离开与 $V_{GS}-V_T$ 的平方律关系， I_D 的影响对n沟比对p沟大。
Λ	有效沟道长度的改变。由于沟道长度的缩短使 I_D 超过 V_{DSat} 时有效沟道长度 L 变短，使得饱和时 I_{Ds} 随 V_{DS} 的增加而增加。
σ	静电反馈系数。这一系数说明 V_T 随 V_{DS} 的增加而下降。 L 减小时 σ 增加。在 V_{GS} 较小时尤其要考虑这一效应。
α	次开启品质因子。它表示电流 I_D 减为1/10所需要减少的栅电压。当 T 增加时， α 也增加。对短沟器件它可能很大。
γ_s	表面体效应参数。当采用开启调整律入时，体效应参数与 V_{SB} 有关。温度增加时， μ_0 和 V_T 减小而 α 增加。
T	电阻。当 I_D 增加时，由于源电阻的存在使有效 V_T 增加。
R	电阻。当 I_D 增加时，由于源电阻的存在使有效 V_T 增加。
V_M	结击穿电压。衬底杂质浓度减小时，结击穿电压增加。
V_{FT}	穿通电压。衬底杂质浓度及 $ V_{SB} $ 增加时，穿通电压增加。

表2.2 各参数值增加时阈值电压的变化

参 数	V_{TN}	V_{TP}
T_{ox}	升	升
N_A 或 N_D	升	升
L	升	升
X_j	降	降
W	降	降
T	降	降
$ V_{DS} $	降	降
$ V_{SB} $	升	升

2.4 互连线的建模

典型的集成电路中占面积最大的不是有源器件而是互连线。在典型的LSI芯片上可能有长达几米的连线，了解它们的特性是很重要的。在本节中，我们将讨论集成电路所用导体的静态和动态性质。

电阻

宽度为 W 、厚度为 T 、长度为 L 的均匀导线的电阻为

$$R = \rho \frac{L}{TW} \quad (2.170)$$

其中 ρ 为材料的电阻率。由于集成电路中导线的厚度在整个芯片范围内可以认为是均匀的，因此，我们可以定义一个新的电阻参数 ρ_s ，称之为薄层电阻，而把 T 吸收在其中。 ρ_s 的量纲为欧姆。我们有*

$$R = \rho_s \frac{L}{W} \quad (2.171)^*$$

铝、重掺杂多晶硅、重掺杂 n^+ 及 p^+ 硅的薄层电阻典型值分别为 0.04, 40, 15 及 25 $\Omega/\square$ 。符号 $\Omega/\square$ 读为欧姆每方。这种有点怪的单位在观察方程 (2.171) 后可以得到解释：所有的正方形在对边之间具有相同电阻。薄层电阻概念特别适合于集成电路还因为它能够很容易地处理沿厚度方向电阻不均匀的问题，这种情况对注入或扩散的导体都会发生。我们有¹⁾

$$\rho_s^{-1} = \int_0^T \rho(x)^{-1} dx \quad (2.172)$$

上面的结果适用于矩形导线。对于复杂形状，必须解 Laplace 方程来确定电阻。这种精确程度对数字集成电路来说几乎是从不需要的，这是因为集成电路的设计至少总要允许有一定的工艺容差。我们要考虑的一种特殊的重要形状是直角拐弯。在直角拐弯中，电阻比按中心线长度计算的要小。把角上的方块计算为 1/2 方计算为一方更准确些。关于形状修正因子的广泛汇集可参阅 Hall [Hall 68, Horowitz 83]。重要的是要根据实际的有效几何尺寸计算电阻而不是只看版图尺寸。对于两臂宽度不同的直角拐弯，为了计算角上矩形的电阻，Bain [Bain 82] 推导了一个近似公式：

$$R_{\text{拐角}} = (0.46 + 0.1a) \rho_s \quad (2.173)$$

其中 a 为宽边与窄边之比。

MOS 集成电路中最常用的三种连线材料为铝、重掺杂多晶硅及重掺杂单晶硅。多晶硅与单晶硅具有相当不同的特性，所以，把它们区别开来是正确的。在 MOS 集成电路中有时采用的材料有金、钨、 WSi_2 及 $MoSi$ 等。采用一种特殊导体的考虑为：可靠性、稳定性、电导率以及与工艺（很可能是高温工艺）的兼容性。

像大多数金属一样，室温下铝的电阻率主要由导电电子与晶格声子的碰撞决定 [Kittel 66]。由于声子数随温度升高而增加，金属的电阻率也随温度升高而增加。室温下纯铝的电阻率约为 $3 \times 10^{-6} \Omega \cdot \text{cm}$ 。在实际的集成电路中，铝用掺有 1—2% 硅的合金铝代替，以减弱铝以尖劈形向硅衬底中钻入的趋势。为了防止金属迁徙往往还加进百分之几的铜。引入这些添加物的考虑将在下一节中讨论。

在单晶硅衬底中的互连线是靠硅表面注入或扩散高浓度的 P, As 或 B 而形成的。连线的电阻率取决于掺的是什么杂质及其分布。杂质原子浓度非常高会使半导体成为简并掺杂，费米能级或是移入导带（对 n^+ ）或是移入价带（对 p^+ ），从而移过整个态密度分布。这使得半导体的特性很像一个金属合金。在这样高的杂质浓度下，电阻率受声子和电离杂质散射二者的限制。电离杂质散射的结果是使室温下硅中电子迁移率由本征值 $1350 \text{ cm}^2/\text{V} \cdot \text{s}$ 下降到约为 $100 \text{ cm}^2/\text{V} \cdot \text{s}$ （当离子化施主浓度为 10^{19} cm^{-3} 时）。重掺杂

* 原文为 $\rho_s = \int_0^T \rho(x) dx$ 有误。——译者注

n^+ 硅的迁移率随温度升高而下降。

硅化物正被日益广泛用于降低芯片上硅及多晶硅连线的电阻。典型硅化物的电阻量为 $2\Omega/\square$ ，并被置于电阻较高的互连线之上。对于短的跨接，接触电阻可能起主要作用，使得硅化物在这些应用中不那么有效。

多晶硅电阻率随杂质浓度的变化与单晶硅材料很不相同。在低杂质浓度下电阻率非常高，在杂质把捕获它们的晶粒间界填满之前，电阻率的下降很慢。超过这一界限后电阻率急剧下降。多晶硅薄膜呈玄武岩结构，晶粒从下面的材料上作柱状生长。晶粒平均尺寸随工艺温度升高而增大。假定晶粒内部电导率与单晶硅相同。晶粒之间的交界包括一些位错原子层，它们的作用是作为不同晶向的晶粒之间的过渡区。关于这些边界的复杂物理机理有几种模型[Lu 81, Seto 75, Colinge 81]。一个主要的效应似乎是载流子被位错所造成的缺陷或不完整的原子键俘获。这种俘获不但减少了可供导电的自由载流子的数目，而且会造成晶粒之间的势垒，从而进一步阻止电流流过。掺杂浓度较低时，多晶硅电阻率随温度升高而下降。但当杂质浓度在 10^{20}cm^{-3} 量级时，温度系数的符号反过来了，变成与单晶硅的符号相同。由于晶粒间界势垒的存在，多晶硅的电流-电压特性实际上略呈非线性。

导体中的功耗和电迁徙

在导体中有两种与电流有关的力作用于晶格原子[Black 69, Chua 81, Heurle 71]。电场对被激活的正离子施加一个力，其方向与电子流的方向相反。由于电子的屏蔽作用，这个力的作用是很小的。作用在晶格原子上的另一个力是被电场驱动的电子流与晶格原子的动量交换，这些碰撞所施加力的方向与电子流的方向相同。这一力的效果是使空位向上游方向移动凝聚成空洞，使离子向下游方向移动形成晶须、小丘。这造成两个问题，首先，空洞的合并会造成连线开路，这一过程是不稳定的，因为空洞的汇集会使该区电流密度增加，反过来，也就会使材料以更大速率移动。失效的第二种模式是由于材料在正端附近汇集，这部分多出的材料会使线间短路。导线的平均失效时间表达式为

$$MTF = \frac{\text{面积}}{\beta J^M} \exp(\phi/kT) \quad (2.174)$$

其中 ϕ 和 β 是取决于材料和晶体结构的常数， M 为 1 到 4 之间的一个常数，通常取为 2， J 为电流密度。波形对称的交流电流引起的迁徙可以忽略。对脉冲电路，电流取 rms 值。 kT 项与激活的离子数有关，因此，高温下电迁徙效应加剧，而在液氮温度下则差不多可以忽略。在采用铝连线的典型集成电路中，为保证电路可靠工作，电流密度须保持在 $1\text{mA}/\mu\text{m}^2$ 以下。观察芯片时把这 $1\text{mA}/\mu\text{m}^2$ 的限制转换为按线宽计的 $\text{mA}/\mu\text{m}$ ，而把金属的厚度考虑在内。 $1\text{mA}/\mu\text{m}^2$ 的数字已包含了一定的安全系数，并已把金属爬过台阶时厚度的减小考虑在内了。金属与多晶硅或扩散区的接触窗口在这方面的问题特别严重，允许通过接触点的电流通常是考虑窗口边长以 $\text{mA}/\mu\text{m}$ 的形式给出，这是因为周界上的薄金属是金属迁徙的限制因素。在导体的抗迁徙能力和它的熔点之间似乎存在着一定的关系。定性地说，这是因为原子在晶格上被固定的越牢，熔点就越高，抗迁徙的能力也越高。

集成电路上另一种主要的电迁徙效应是由于硅熔入铝中而引起的。溶入的硅很容易

被电子流从负的接触端搬走，使得铝钻入消失的硅接触中。最后，这使得与源区相连的铝钻透到衬底中。为了防止这种钻透，有时在Al与Si之间放上一层阻挡金属如Pt，这同时也降低了接触电阻。

导体上单位长度的功耗为 $\rho_s I^2 / W$ 。对金属和多晶硅层来说，导体上的热必须通过 SiO_2 绝缘层传出去。 SiO_2 是不良的热导体，比硅约差 100 倍。导体的温升近似为

$$\Delta T = \frac{\rho_s I^2 H}{W^2 K_{\text{thermal}} - \text{SiO}_2} \quad (2.175)$$

其中 H 为衬底上面导线的高度， K_{thermal} 为热导率。例如，若 $H = 0.6 \mu\text{m}$ ， $W = 2 \mu\text{m}$ ， $\rho_s = 40 \Omega / \square$ ， $I = 25 \text{mA}$ ，则有 $\Delta T = 2700^\circ\text{C}$ 。聚酰亚胺是一种有机电介质，其热阻比二氧化硅的热阻大 10 倍。

互连线电容

信号互连线网络的电容，有时还有其电阻，在逻辑门的延迟中起主要作用。不但平行平板对电容有重要贡献而且边缘电场对窄线电容也有重要贡献。对于 $2 \mu\text{m}$ 的工艺，金属或多晶线的边缘电容可能与平行平板电容一样大。

一条导线与接地平面或另一条导线之间的电容是通过解 Laplace 方程确定的。我们将考察两个有已知解的特殊情形而不再去解这一电场问题。一个特殊情形是平行板电容。电容为

$$C_{\text{平板}} = \frac{\epsilon W L}{H} \quad (2.176)$$

电容与面积 WL 及介电常数 ϵ 成正比，与平板之间的距离 H 成反比。当 $W, L \gg H$ 时， $C_{\text{平板}}$ 与真实电容很接近。

第二个特殊情形用于窄导线，即 $L \gg H$ ， $W \leq H$ 。如果我们把导线近似为长圆柱，则可得到电容的简单表达式：

$$C_{\text{圆柱}} = \frac{2\pi\epsilon L}{\ln \left[1 + \frac{2H}{W} \left(1 + \sqrt{1 + \frac{W}{H}} \right) \right]} \quad (2.177)$$

注意圆柱电容与 W 及 H 的对数有关。 W 可解释为圆柱的直径。观察这两个特殊情形对按比例缩小互连线可得到一个重要启示。对于给定的绝缘体厚度 H ，通过改进复印和刻蚀技术缩小连线宽度，只有当线宽尚未小到与绝缘体厚度同数量级时，电容才是线性减小的。在此之后，只能以对数率作进一步改进。

典型互连线的电容是用平行板电容公式计算的电容与边缘场电容的某种组合。可以用方程 (2.176) 和 (2.177) 来近似计算一根导线的总电容。我们定义

$$C_{\text{等效}} = C_{\text{平板}} + 2C_{\text{边缘}} \quad (2.178)$$

其中 $C_{\text{平板}}$ 由方程 (2.176) 给出。 $C_{\text{边缘}}$ 可以近似如图 2.39 所示。我们用半圆柱来代替导线的侧面。这样，边缘电容就可以近似为圆柱电容的一半再减去平行板电容的一小部分。我们有

$$C_{\text{边缘}} \approx \epsilon L \left\{ \ln \left[1 + \frac{2H}{T} \left(1 + \sqrt{1 + \frac{T}{H}} \right) \right] - \frac{T}{4H} \right\} \quad (2.179)$$

对于 $H = T$ 、 $\epsilon = 3.9\epsilon_0$ 的情形， $C_{\text{边缘}} \approx 0.044 \text{ fF}/\mu\text{m}$ ，这一数字基本与 T 无关。

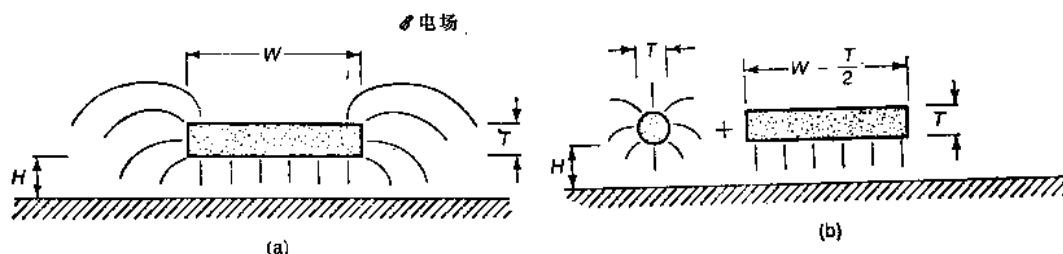


图2.39 窄线的边缘电容。(a) 导线的边缘电场电力线；(b) 把导线看作一个平行板电容器加上接地平面上的柱状导体电容的模型

对于间距与其线宽差不多的导线，一根导线与相邻两根导线之间的耦合电容与它对接对平面 (W/H 在 1 到 2 之间) 的电容差不多相等 [Dang 81]。Lewis 给出了 VLSI 中导线耦合电容的公式 [Lewis 84]。这种线间电容在考虑噪声时是非常重要的。

方程 (2.179) 适用于多晶硅及金属线。扩散或注入单晶硅形成的互连线，其边缘电容更大。这是因为场注入使导线侧面的耗尽层更薄，这一耗尽层起着绝缘体的作用。这种侧壁耗尽层电容往往成为相邻互连线间电容的主要成分。耗尽层电容是非线性的，其表达式为

$$C = \frac{C_J W L}{\left(1 + \frac{V}{\phi_j}\right)^{M_J}} + \frac{2C_{JSW}(L+W)}{\left(1 + \frac{V}{\phi_{JSW}}\right)^{M_{JSW}}} \quad (2.180)$$

其中 V 为导线对衬底的电压， ϕ 为结的自建势， M 为一个常数，对缓变结约为 0.3，对突变结为 0.5。交叉或交叠线之间的电容可以用上面提到的方法近似计算，也可以参阅微波带状线或微带资料得到更精确的结果。

例2.8 nMOS超缓冲器寄生量的计算。

要开发如图1.32所示的超缓冲器电路模型就牵涉到电容的计算，在本例中，我们将忽略电阻的影响。由此得出的电路图如图2.40所示。结点既标出了它们的名字也标出了

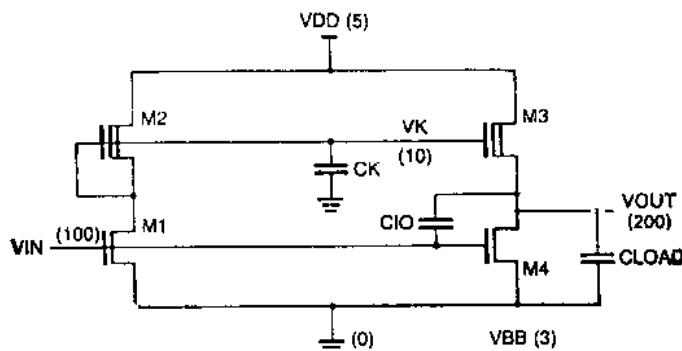


图2.40 超缓冲器的模型。括号中所示为SPICE结点编号

它们在SPICE中的结点编号。地永远标为结点0。

晶体管参数，包括栅电容、交叠电容、源/漏电容，将是晶体管模型的一部分。后面我们将讨论这些。现在我们先看看四个固定电容CIN，CK，CIO，及CLOAD。CIN由几部分组成，包括金属线对场区的电容 C_{mf} 及边缘电容，多晶对场区电容 C_{polyf} ，金属对多晶电容 C_{mpoly} 等。边缘电容假定为 $0.044\text{fF}/\mu\text{m}$ 。所有其它值都可由附录C中取出，它们都取最大（最坏情形）值。注意在输入接触处对金属-多晶电容没有加以计算，这是因为这两条线是互连在一起的。金属对多晶的电容出现在下拉晶体管区。在输入多晶线与输出金属线交叉处有一个密勒电容CIO。我们略去了多晶界边缘电容。表2.3中列出了这些固定电容值。

表2.3 固定电容

类 型	CIN	CK	CIO	CLOAD
C_{mf} (场区)($\mu\text{m}^2/\text{pF}$)	200/0.0084	0/0	0/0	154/0.0069
C_{mf} (边缘)($\mu\text{m}/\text{pF}$)	100/0.0044	0/0	0/0	98/0.0043
C_{mpoly} ($\mu\text{m}^2/\text{pF}$)	16/0.0018	48/0.0053	8/0.0009	0/0
C_{polyf} ($\mu\text{m}^2/\text{pF}$)	48/0.0034	44/0.0031	0/0	0/0
C_{mdiff} ($\mu\text{m}^2/\text{pF}$)	0/0	0/0	0/0	56/0.0062
C_d (pF)	0.0180	0.084	0.0009	0.0174

由版图提取的晶体管参数在进入SPICE之前需要就 ΔL_{poly} 和 ΔW 作一调整。最坏情形 ΔL_{poly} 为 $-0.125\mu\text{m}$ ，最坏情形 ΔW 为 $0.9\mu\text{m}$ （对长的窄沟晶体管）。在SPICE中扩散区电容是通过面积（AD和AS）和周界（PD和PS）输入的。我们把紧挨着晶体管的扩散区长度都计入扩散边界（PD或PS）。这并不很正确，但可作为合理的近似。对于埋孔接触我们也没有作任何特殊处理，尽管由于掺杂P与As的不同，埋孔接触的结比源/漏的结要深。同样，这也是一种合理的近似。关于这样的处理，对埋孔接触我们乐观一些，而对晶体管边界电容则悲观一些。

我们对下拉晶体管的漏区和上拉晶体管的源区扩散电容的区分是十分任意的，只要所有各电容都已计算在内就行了。显然，我们不需要计算地线和 V_{DD} 的电容。表2.4所示为最后的晶体管参数。表2.5所示为超缓冲器的SPICE输入文件。

表2.4 晶体管参数

	M 1	M 2	M 3	M 4
版图W/L	13/2	6/6	15/6	23/2
SPICEW/L	11.2/2.25	4.2/6.25	13.2/6.25	21.2/2.25
AD	61	0	0	0
AS	0	0	110	0
PD	42	0	0	0
PS	0	0	62	0

表2.5 SPICE输入文件

SPICE DECK for Fig. 2.40
* MODEL NENHS NMOS LEVEL=3 RSH=0 TOX=330E-10 LC=0.19E-6 XJ=0.27E-6 + VMAX=13E4 ETA=0.25 KAPPA=0.5 NSUB=5E14

```

UO=650 THETA=0.1
* VTO=0.946 CGSO=2.43E-10 CGDO=2.43E-10
CJ=6.9E-5 CJSW=3.3E-10
+ PB=0.7 MI=0.5 MISW=0.3 NFS=1E10
*
*****
VIN 100 0 PULSE (5 0 6NS 2NS)
VDD 5 0 5
VBB 3 0 -3
OPTIONS DEFL=2.25E-6
WIDTH IN=75 OUT=75
* node 10 is VK
CIN 100 0 0.018PF
CK 10 0 0.0084PF
CIO 10 200 0.0009PF
CLOAD 200 0 0.0174PF
CHACK 200 0 1PF
* CHACK is an additional output load to make things more realistic
M1 10 100 0 3 NENHS W=11.2U AD=61P PD=42U
M2 5 10 10 3 NDEPS W=4.2U L=8.25U
M3 5 10 200 3 NDEPS W=13.2U L=6.25U AS=110P PS=62U
M4 200 100 0 3 NENHS W=21.2U
* TRAN 0.3NS 15NS
* PLOT TRAN V(200) V(10) V(100) (0,6)
* END

```

RC线

在MOS集成电路中常有RC梯形和树形结构。图2.41所示就是一个RC梯形结构。长的多晶或扩散线以及一串传输晶体管都是其中一例。这种网络可以是非线性的或是分布的，或二者兼而有之。本节中我们将讨论线性RC梯形网络的动态特性。通过讨论这种特殊的但却是重要的情形，我们可以对这类电路的突出特点有个很好的了解，同时也可以是对有关文献的入门。利用这种形式的网络也可以模拟热的流动和扩散。

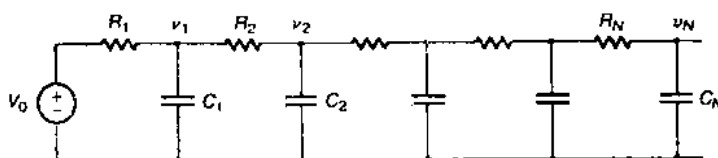


图2.41 RC梯形网络 (v_N 是输出节点)

我们首先讨论由 N 个电阻和 N 个电容组成的梯形网络对阶跃电压的响应。输入电压从零升到 V_0 。第 n 个结点上的电压可写为

$$v_N(t) = V_0 - \sum_{j=1}^N C_j \frac{dv_j}{dt} \sum_{i=1}^j R_i \quad (2.181)$$

其中 $C_i \frac{dv_i}{dt}$ 为通过与第 j 个结点相连的电容流到地的电流，它与 $\sum_{i=1}^j R_i$ 相乘*代表由该电流产生的压降。外层的求和是所有这些压降的叠加。整理方程 (2.181) 并对等式两边从 $t = 0$ 到 $t = T_D$ 积分，可得

$$\int_0^{T_D} [V_0 - v_N(t)] dt = \sum_{j=1}^N \sum_{i=1}^j R_i C_i v_i \quad (2.182)$$

T_D 为第 N 个结点电压达到 v_N 的时间。我们已假定初始条件为所有结点电压为零。可以指出 [Protonotarios 67]

$$v_N \leq v_i \leq V_0 \quad \text{对所有 } i \quad (2.183)$$

这就是说，第 j 个结点的电压永远小于等于 v_{j-1} ，我们定义一个常数 T_E ：

$$T_E \equiv \sum_{j=1}^N C_j \sum_{i=1}^j R_i \quad (2.184)$$

由式 (2.183) 可得

$$v_N T_E \leq \sum_{j=1}^N \sum_{i=1}^j R_i C_i v_i \leq V_0 T_E \quad (2.185)$$

函数 $v_N(t)$ 为随时间增加的单调函数。因此，我们也可得出

$$T_D(V_0 - v_N) \leq \int_0^{T_D} [V_0 - v_N(t)] dt \leq T_D V_0 \quad (2.186)$$

对于给定的输出电压 v_N ， T_D 为输出达到这一电压所需时间。利用方程 (2.182)，(2.185) 和 (2.186) 我们可得出对这一时间的限制条件，即

$$\frac{v_N}{V_0} T_E \leq T_D \leq \frac{V_0}{V_0 - v_N} T_E \quad (2.187)$$

看看一种特殊情形是有帮助的。假定所有电阻都为 R_0 ，所有电容都为 C_0 ，则

$$T_E = R_0 C_0 \frac{N(N+1)}{2} \quad (2.188)$$

我们可以看出均匀 RC 梯形网络的延迟与元件数目的平方成正比。从物理上说，这意味着 RC 线的延迟与其长度的平方成正比。这一效应从概念上也可以这样来理解：延迟正比于总电容乘总电阻再乘一个近似系数，该系数取决于 v_N/V_0 及具体电路结构。总电容及总电阻二者都与延迟线长度成正比。若 $N \rightarrow \infty$ 时，则 $T_E \rightarrow R_L C_L / 2$ ，其中 R_L 为总的线电阻， C_L 为总的线电容。Rubinstein 等人求出了比方程 (2.187) 更严格的限制。他们指出，除了方程 (2.187) 之外，下式也成立 [Rubinstein 83]：

$$T_D \geq T_E - T_R \left\{ 1 + \ln \left[\frac{T_E}{T_R} \left(1 - \frac{v_N}{V_0} \right) \right] \right\} \quad (2.189)$$

$$T_D \leq \frac{T_E}{1 - \frac{v_N}{V_0}} - T_R \quad (2.190)$$

$$T_D \leq T_E \left[1 - \ln \left(1 - \frac{v_N}{V_0} \right) \right] - T_R \quad (2.191)$$

* 原文为 $\sum_{i=1}^j R_i$ 代表……，有误。——译者注

其中

$$T_R = \frac{\sum_{j=1}^N C_j \left(\sum_{i=1}^j R_i \right)^2}{\sum_{j=1}^N R_j} \quad (2.192)$$

所有这些限制条件经过变换后, 可以给出在给定延迟 T_D 下对电压 v_N 的限制条件。

例2.9 ROM的SPICE模型。

现有一ROM阵列, 它有256行字线和256列位线。字线为多晶硅, 位线为金金属。线分布在 n^+ 区, 每隔15根位线有一根与金属地线相连。其版图如图2.42所示。对这一矩阵我们开发了一个模型来帮助预测最坏情形速度。

沿多晶字线的 RC 延迟是相当大的。为了避免模拟256个线段, 我们把一条长线分割成数目较少的线段。8个线段是个合理的数目, 这一合理性可以通过对10个线段的模拟得到证实, 模拟表明, 其波形与8个线段的情形没有多大区别。设多晶字线每一位长度的电阻为 R_{word} , 则每一线段的电阻为

$$32R_{word} + \frac{32}{15}R_{refresh}$$

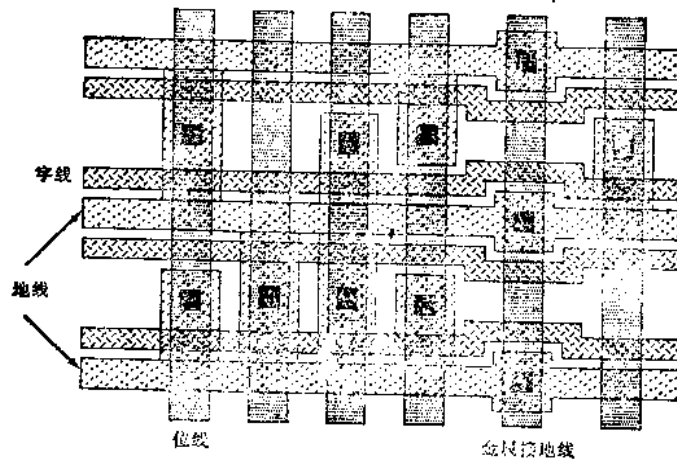


图2.42 ROM单元版图

其中 $R_{refresh}$ 为多晶在穿过接地金属线区的电阻, 电容有4个组成部分。第一部分是多晶对固定电位的电容, 称之为 C_{word} 。第二部分是金属位线的耦合电容 C_c 。第三部分为晶体管的栅电容。第四部分是由于版图上有接地金属线而产生的一个小电容 $C_{refresh}$ 。假定字线上挂满晶体管, 该字线的模型如图2.43所示。注意位线电压 v_{bit} 是模型的一部分。后面我们将要讨论产生 v_{bit} 的方法。

由于问题相当复杂, 我们的目标是产生一个模型装入SPICE。所以我们并不计算晶体管的输入电容, 而把这留给SPICE去做。我们所做的一个近似是假定所有位线的电压是一样的。对于最坏情形延迟, 所有位线的电压将是下降的, 但它们并不正好在同一时

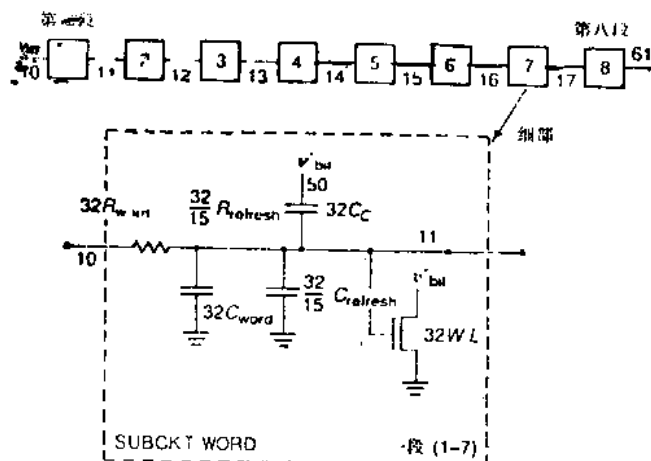


图2.43 字线模型。头七段是一样的，第八段则有区别

间下降，这是因为要考虑多晶的 RC 延迟。不过，假定它们在同一时间下降也并不是很差的近似，因为延迟的 C_c 分量并不那样大。图 2.43 中的模型将被用于八个线段中的七个。第八段的模型需要更复杂些。

建模的下一步涉及位线及如何产生 v_{bl} 。最慢的位线是在矩阵远端的两条地线之间的位线(第249根位线或其左右)。 R_s 为每个单元的地线电阻。在数方块以计算 R_s 时需要考虑 ΔW 的影响。在最坏情形，所有15个晶体管都是导通的。我们希望能找到不去模拟全部15个晶体管的方法。由于对称性，在第8个晶体管左面和右面流过的电流是一样的。从概念上说，如果我们把等电位的结点短接起来，我们可以用图2.44所示网络模拟下拉过程。7个晶体管和7个电阻并排地连接起来。 R_c 为接触电阻。由于它很小(见表2.6)，我们将予以忽略。图2.45所示为实际下拉位线的单元模型。压控电压源被用来产生 v'_{bl} 和 v'_{word} 而不考虑把所有位线、字线负载都加上去。 v'_{bl} 用于所有线段。

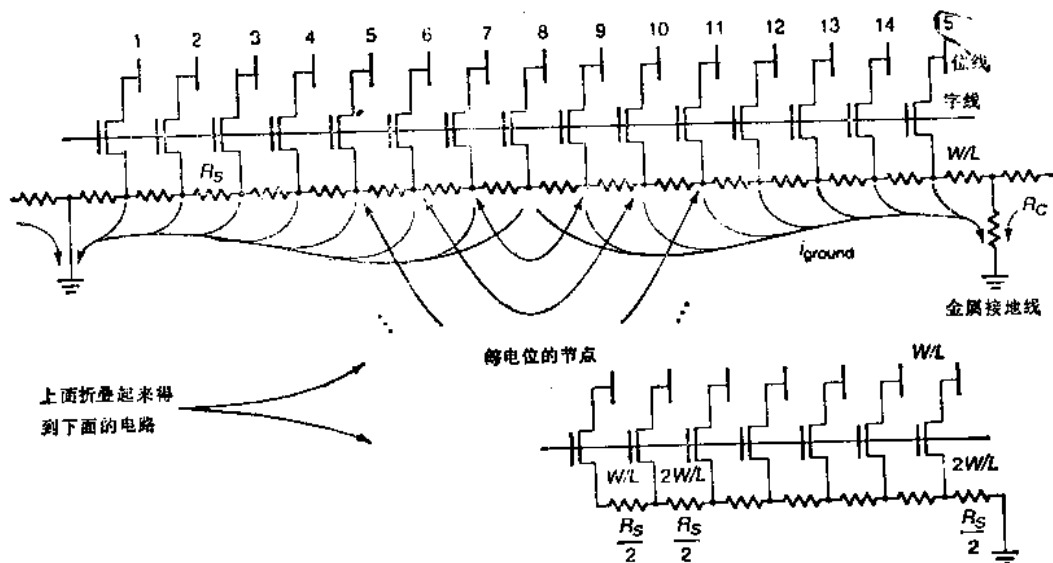


图2.44 在金属接地线模型中，利用对称性来减少元件的数目，等电位的节点可以短接在一起

表2.6 模拟模型参数

参数	数值	单位	注 释
R_C	4	Ω	不重要
R_{word}	142	Ω	$40 \times 8 / 2.25$
R_s	258	Ω	5×50
$R_{refresh}$	213	Ω	5.3×40
W	4.2	μm	$6 - 2 \times 0.9$
L	2.25	μm	$\Delta L = -0.125 \mu m$
C_{bit}	11×10^{-4}	pF	$C_{mdiff} = 11 \times 10^{-6} pF / \mu m^2$
C_{word}	2.8×10^{-4}	pF	$C_{polyf} = 7 \times 10^{-6} pF / \mu m^2$
C_C	8.8×10^{-4}	pF	$C_{mpoly} = 11 \times 10^{-6} pF / \mu m^2$
$C_{refresh}$	3.12×10^{-8}	pF	$32 \mu m^2 C_{polyf} + 8 \mu m^2 C_{mpoly}$
AD	4920	μm^2	$52 \times 42 + 76 \times 36$
PD	3176	μm	$52 \times 26 + 76 \times 24$

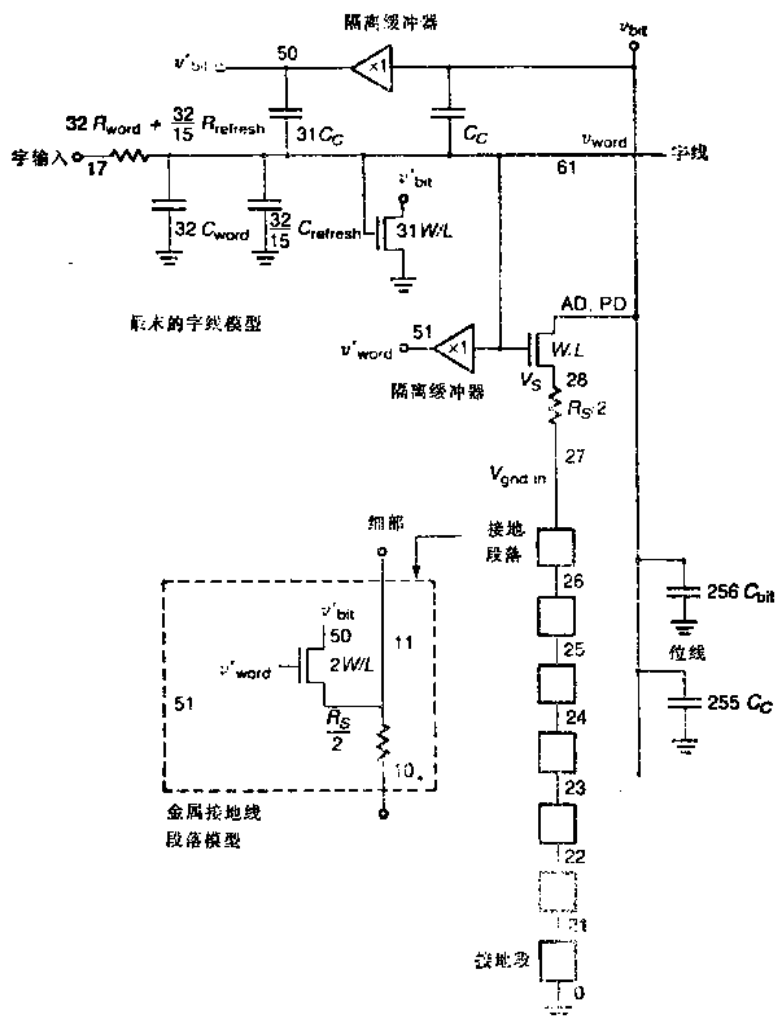


图2.45 ROM模型的第八段。压控电压源在没有电容负载的条件下被用于产生其它段所需要的电压

关于晶体管，我们没有提出位线上漏电容的模型，而把面积和周界送入SPICE。假定在最坏情形下位线上70%的位置有晶体管。这180个晶体管最坏的安排是开所有128个源接触孔，其中52个接触孔连接两个晶体管，76个孔只连一个晶体管。

现在我们用附录中 $2\mu\text{m}$ nMOS工艺的实例来计算这些参数的实际值。我们感兴趣的是最慢的情形，所以我们采用最大的寄生电容值和多晶电阻最大值¹⁾。

对模型的各种参数求解后，我们得出表2.6中所给出的值。

对某些有奇怪几何形状的情形，计数方块可能是困难的。在这些情形中，追求超过参数本身重要性的精确度是不可取的。图2.46至图2.53说明了如何从版图提取各种参数。

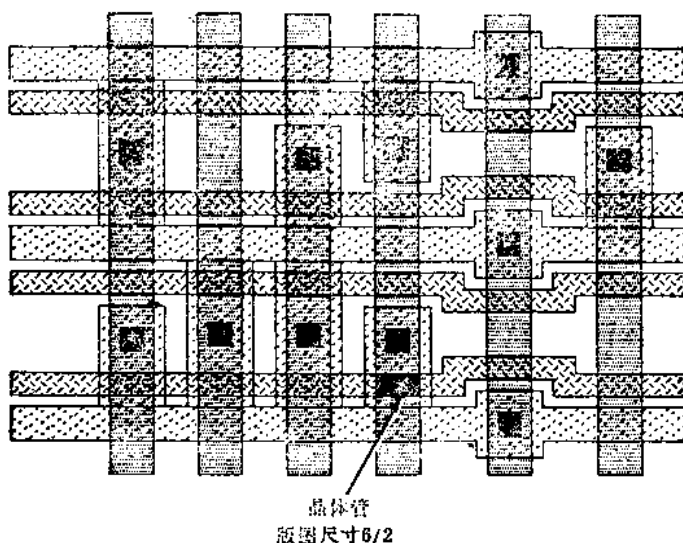


图2.46 为SPICE计算晶体管的宽度和长度， $W = 6\mu\text{m} - 2\Delta W$ ， $L = 2\mu\text{m} - 2\Delta L_{\text{poly}}$

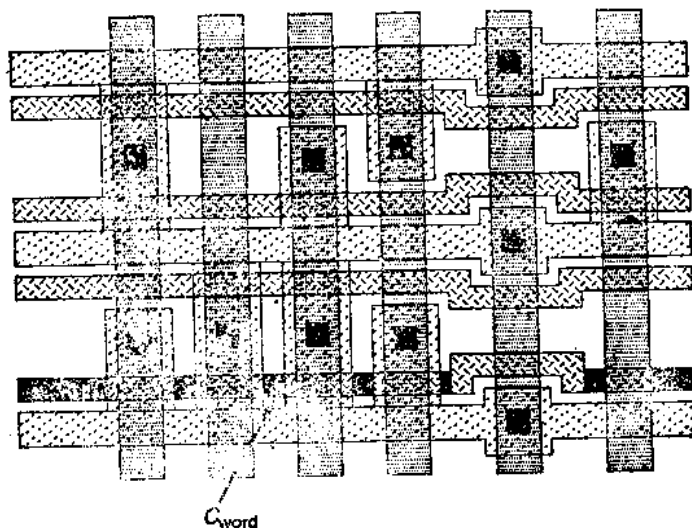


图2.47 字线电容 $C_{\text{word}} = 4\mu\text{m}^2 \times C_{\text{poly}}$

1) (在读了第四章后读此注) 对位线模型，最坏情形下晶体管 ΔL_{poly} 取负值，而多晶 RC 最坏情形实际上在 ΔL_{poly} 为正的时候才发生，如果我们将边缘电容考虑在内。这种交叉过程的模拟需要调整“慢”工艺文件。

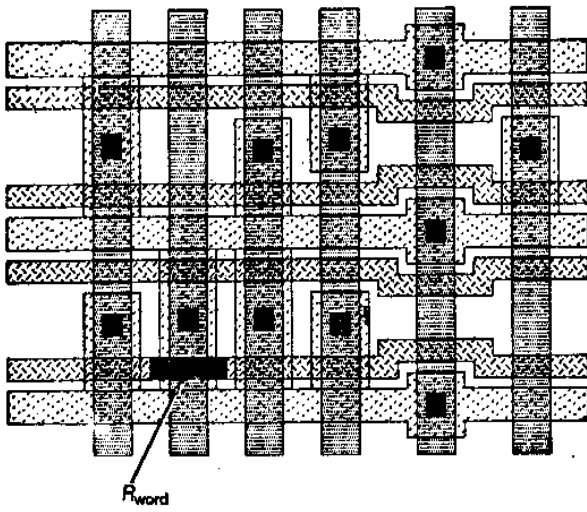


图2.48 字线电阻。 $R_{word} = (8\mu m \times R_{po\gamma}) / (24m - 2\Delta L_{poly})$

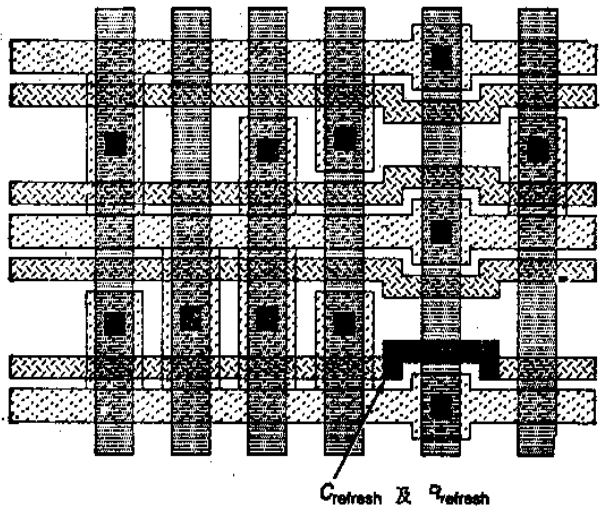


图2.49 接地段电容及电阻。 $C_{refresh} = 32\mu m^2 \Delta L_{po\gamma} + 8\mu m^2 C_{mpoly}$ $R_{refresh} = R_{poly} (16\mu m - 4\mu m) / (24m - 2\Delta L_{po\gamma})$ 。计算中我们把角上的电阻按半方计算

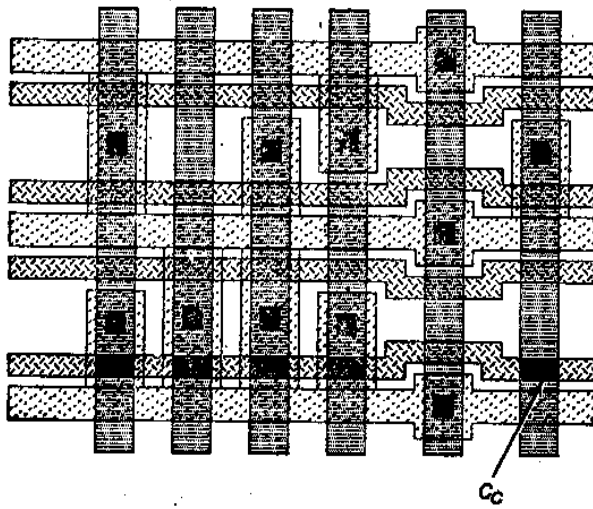


图2.50 金属对多晶的耦合电容。 $C_c = 8\mu m^2 C_{mpoly}$

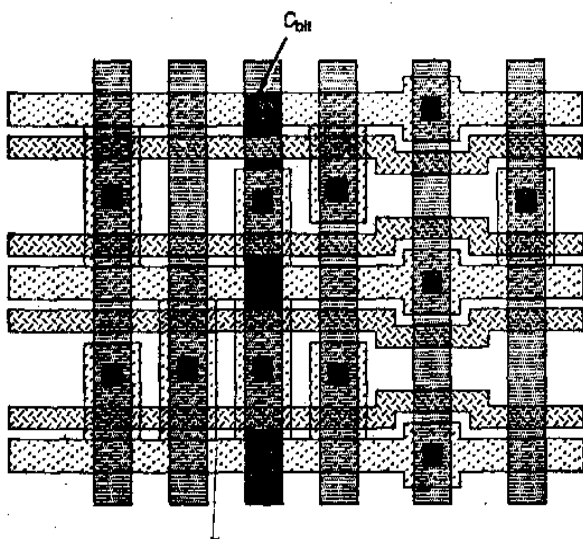


图2.51 金属位线电容主要是n⁺上金属的电容，略去了场区上金属的电容。 $C_{bit} = 20\mu\text{m}^2 C_{mdiff}/2$

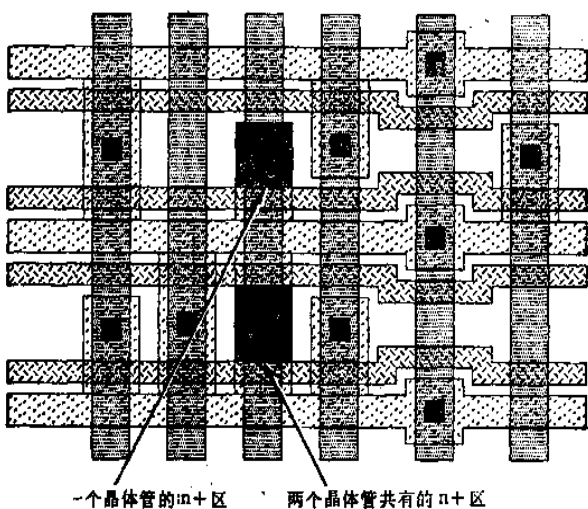


图2.52 在SPICE中扩散电容是用栅区面积和周界输入的。在双晶体管位置， $AD = 42\mu\text{m}^2$ ， $PD = 26\mu\text{m}$ 在单晶体管位置， $AD = 36\mu\text{m}^2$ ， $PD = 24\mu\text{m}$

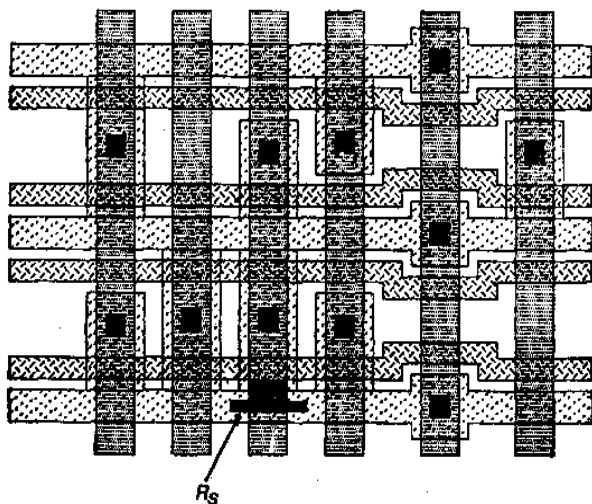


图2.53 扩散地线的串联电阻， $R_g \approx R_{diff}(2R_{tab} + R_{middle})$ 。其中， $R_{tab} = (1\mu\text{m} + \Delta W)/(3\mu\text{m} - 2\Delta W)$ ， $R_{middle} = (6\mu\text{m} - 2\Delta W)/(3\mu\text{m} - 2\Delta W)$

计算 C_{bit} 时，我们只考虑金属对接地 n^+ 区的电容。金属对场区也有一点电容但不考虑。每个字有 $10\mu m^2$ 的金属叠在 n^+ 区之上。在 $C_{refresh}$ 中，我们既考虑多晶对场区电容也考虑多晶对其上的接地金属线的电容。

表2.7给出了一个SPICE输入卡，图2.54所示为SPICE模拟输出。请注意字线延迟与地线噪声。

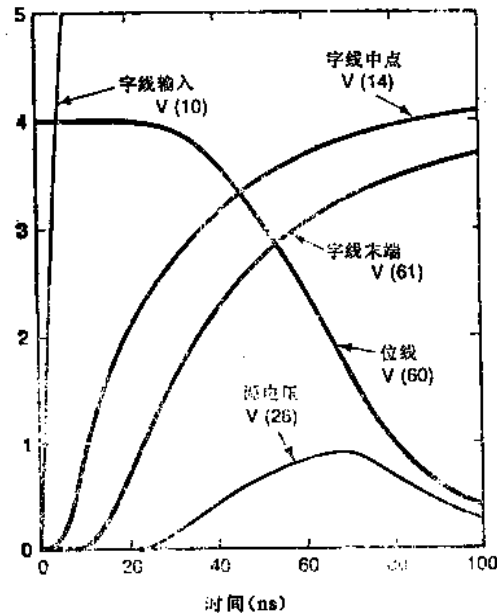


图2.54 模拟的ROM波形。注意观察下拉器件的源电压，它的上升对速度有严重影响

表2.7 ROM模拟的SPICE卡

```
ROM simulation model (ROM模拟模型)
*
* node names worth remembering (重要节点名字)
* VBIT 60
* VBIT' 50
* VWORD 61
* VWORD' 51
*
VBB 2 0 -3V
*
* step input (阶跃输入)
VIN 10 0 PWL (ONS 0 6NS5500NS 5)
*
*
* word line section model (字线段模型)
* VIN VOUT VBIT' VBB
SUBCKT WORD 10 11 50 2
RWORD 10 11 5006 2
CWORD 11 0 0.009PF
* CRFSH=32/15 C-REFRESH
CRFSH 11 0 0.0067PF
* 32 cells worth of CC (32个单元的CC)
CC 50 11 0.0282PF
```

```

* no need for AD, PD since VBIT' is driven from VCCS (由于VBIT'是由VCCS驱动的, 所以不需要
AD, PD)
MWORD 50 11 0 2 NENHS W=134.4U L=2.25U
.ENDS WORD
*
* rish line section model (金属接地线段模型)
* VIN VOUT VWORD' VBIT' VBB
.SUBCKT RFSH 10 11 51 50 2
* RRFSH=RS/2
RRFSH 10 11 129.17
MRFSH 50 51 11 2 NENHS W=8.4U L=2.25U
.ENDS RFSH
*
* the whole thing (全部文件)
*
* 7 word line sections (7个字线段)
X1 10 11 50 2 WORD
X2 11 12 50 2 WORD
X3 12 13 50 2 WORD
X4 13 14 50 2 WORD
X5 14 15 50 2 WORD
X6 15 16 50 2 WORD
X7 16 17 50 2 WORD
* the last word line section is different (字线最后一段不同)
RWORD 17 61 5006.2
CWORD 61 0 0.009PF
* CRFSH=32/15 C-REFRESH
CRFSH 61 0 0.0067PF
* 31 cells worth of CC (31个单元的CC值)
CCS 50 61 0.0273pF
CC 60 61 0.0009PF
MWORD 50 61 0 2 NENHS W=130.2U L=2.25U
*
* 7 refresh sections (7个接地线段)
X11 0 21 51 50 2 RFSH
X12 21 22 51 50 2 RFSH
X13 22 23 51 50 2 RFSH
X14 23 24 51 50 2 RFSH
X15 24 25 51 50 2 RFSH
X16 25 26 51 50 2 RFSH
X17 26 27 51 50 2 RFSH
* the last refresh section is different (W/L vs 2W/L) (接地线段是不同的, (W/L对2W/L))
RRFSH 27 28 129.17
MRFSH 60 61 28 2 NENHS W=4.2U L=2.25U AD=4920 OP PD=3176 OU
CBIT 60 0 0.2816PF
CCBIT 60 0 0.2244PF
*
* unity gain isolation buffers (增益为1的隔离缓冲器)
EBIT 50 0 60 0 1.0
EWORD 51 0 61 0 1.0

```

```

*
* precharge the bit line (位线预充电)
.IC V(60) = 4
*
*
* TRAN 1NS 100NS
* PRINT TRAN V(10) V(14) V(28) V(60) V(61)
* .PLOT TRAN V(10) V(14) V(28) V(60) V(61)
* OPTION LIMPTS=1001
*
*
* *****
* worst case slow parameters (最坏情形的慢参数)
*
* .MODEL NENMS NMOS LEVEL=3 RSH=0 TOX=330E-10 LD=0.19E-6 XJ=0.27E-6
+ VMAX=13E4 ETA=0.25 KAPPA=0.5 NSUB=5E14 UO=650 THETA=0.1
+ VTO=0.946 CGSO=2.43E-10 CGDO=2.43E-10 CJ=6.9E-5 CJSW=3.3E-10
+ PB=0.7 MJ=0.5 MJSW=0.3 NFS=1E10
*
* .MODEL NZEROS NMOS LEVEL=3 RSH=0 TOX=330E-10 LD=0.19E-6 XJ=0.27E-6
+ VMAX=13E4 ETA=0.25 KAPPA=0.5 NSUB=4.0E14 UO=680 THETA=0.1
+ VTO=0.526 CGSO=2.43E-10 CGDO=2.43E-10 CJ=6.9E-5 CJSW=3.3E-10
+ PB=0.7 MJ=0.5 MJSW=0.3 NFS=1E10
*
* .MODEL NDEPS NMOS LEVEL=3 RSH=0 TOX=330E-10 LD=0.19E-6 XJ=0.27E-6
+ VMAX=13E4 ETA=0.25 KAPPA=0.5 NSUB=50E14 UO=650 THETA=0.04
+ VTO=-2.078 CGSO=2.43E-10 CGDO=2.43E-10 CJ=6.9E-5 CJSW=3.3E-10
+ PB=0.7 MJ=0.5 MJSW=0.3 NFS=1E10
*
* .END

```

2.5 散 热

VLSI产生的热位于芯片内部不易散去。如果在设计nMOS VLSI时,对芯片功耗不特别注意节制则门平均功耗很容易在1mW的量级。将这一值乘以VLSI芯片上的门数,其功耗将是非常大的。如Keyes所指出(Keyes 75),日光的热量大约只有 $0.1\text{W}/\text{cm}^2$,可以把物体照得相当热;一个表面积为 120cm^2 的60W灯泡,其功耗仅为 $0.5\text{W}/\text{cm}^2$,就热得不能摸了,那么,可以想象,从每边只有几个毫米的芯片上耗散几瓦的功率该是怎样一个问题了。

芯片功耗的代价在功耗为1W以下时主要归结为高质量电源的价格。在1W以上,冷却成本将成为主要因素。当芯片功耗超过1W时,所需的特殊封装、风扇、及致冷设备等使得与功耗相联系的代价迅速上升。对目前的封装,实际的最高极限值可能是每片8W左右。功耗能达到 $25\text{W}/\text{cm}^2$ 的封装正处在规划阶段。近千瓦的功率已成功地从硅芯片上散走,办法是在硅片背面刻一些槽,并强使液体通过以达到极好的热交换效果。

要从物体上散热,在物体与散热器之间必须有温度差。静止空气可以散的热约为

$0.001\text{W}/\text{cm}^2\cdot^\circ\text{C}$ 。所以,设封装表面积为 10cm^2 ,则它与空气的温度差约为 $100^\circ\text{C}/\text{W}$ 。利用强制通风,可以耗散的功率升高到 0.01 到 $0.03\text{W}/\text{cm}^2\cdot^\circ\text{C}$ 之间。对于这样低的耗散功率,硅片与封装几乎为相同温度。

如果功率的耗散通过空气冷却不够,可用液体冷却。当热源与液体的温度差不大时,传热主要靠对流。例如氟利昂的沸点为 48°C ,在温度差为 20°C 时可传热为 $20\text{W}/\text{cm}^2$ 。当温度差较大时,发生有核沸腾。最终,沸腾变得如此激烈以致在热源与液体之间形成一层蒸气,成为一个阻挡层阻碍进一步的有效散热。

芯片的最高温度受两个因素限制,即可靠性及允许的晶体管参数(首要的是迁移率及阈电压)的最大退化。从可靠性说,塑封芯片的最高温度约为 100°C ,对陶瓷封装芯片为 120°C 。封装好的芯片的实际温度可以通过测量一个经校准过的 p-n 二极管的脉冲 $I-V$ 移性来测定。采用脉冲方法是为了防止加热芯片。把器件放在烘箱中,在给定电流下测量其前向电压,由此可以校准电压与二极管温度的关系。然后,当芯片工作时,可以再一次进行这种电压测量,通过校准曲线查出二极管的温度。

我们定义 θ_{ja} 为结与外界环境之间的热阻。对大多数商业封装,这基本上就是封装的热阻。热阻 θ_{ja} 为芯片功耗 1W 时芯片与外界静止空气之间的温度差($^\circ\text{C}$)。常用封装的热阻约在 $30-70^\circ\text{C}/\text{W}$ 之间。双列直插(DIP)40脚塑料封装的 θ_{ja} 约为 $62^\circ\text{C}/\text{W}$,而64脚陶瓷封装的热阻约为 $36^\circ\text{C}/\text{W}$ [Murtuza 82]。这些热阻都受封装的环境的影响。所以印制电路板的位置和元件安排都会影响封装的有效热阻。相当一部分热可以从引线散出去。在强制通风下($18\text{m}^3/\text{min}$), θ_{ja} 可以下降 $10-15^\circ\text{C}/\text{W}$ 。对于一个不太小的芯片,与它封装之间的热阻约在 $3^\circ\text{C}/\text{W}$ 的量级,对空气冷却的封装,这通常是可以忽略的。另一方面,若封装得不好,其热阻可能大得多,这在任何情况下都是个大问题。

近来在芯片的封装技术方面进行了大量的研究与开发。人们可以期待,经过几年会有迅速的进展。当然,由于与功耗有关的成本具有类似阈值的特点,使得确定功耗的限额很直接了当。当然,在这些制约条件下进行设计,就是另外一回事了。

2.6 小 结

器件电子学是VLSI电路设计的基础。MOS器件物理和其外特性之间有密切的关系。MOS晶体管是一个电压控制的非线性电阻。有无数现象影响 $V-I$ 特性的各个方面,但绝大多数效应都可以通过电荷分布及载流子速度来解释。一个关键参数是饱和电流,它限制着通过晶体管向电容充电的速度,另一个关键参数是阈电压。我们已了解到阈电压不是一个常数,它随所加电压及晶体管尺寸而改变。图2.55所示为模拟所得的 $1\mu\text{m}$ nMOS晶体管掺杂浓度分布。图2.56所示为这一器件饱和状态下的静电电位分布。图2.57及2.58为电子与电流密度的分布。注意载流子数目沿着沟道从源到漏减少的情况。图2.59所示为推动载流子的侧向电场,注意它在沟道的漏端附近达到峰值。

MOS晶体管的电容主要取决于栅氧化层,但是我们也了解另外两个重要效应。变电容效应使MOS器件的输入电容在阈电压以上急剧增加。MOS晶体管饱和时,栅与漏的电容极小。这对设计耗尽型负载是十分重要的。

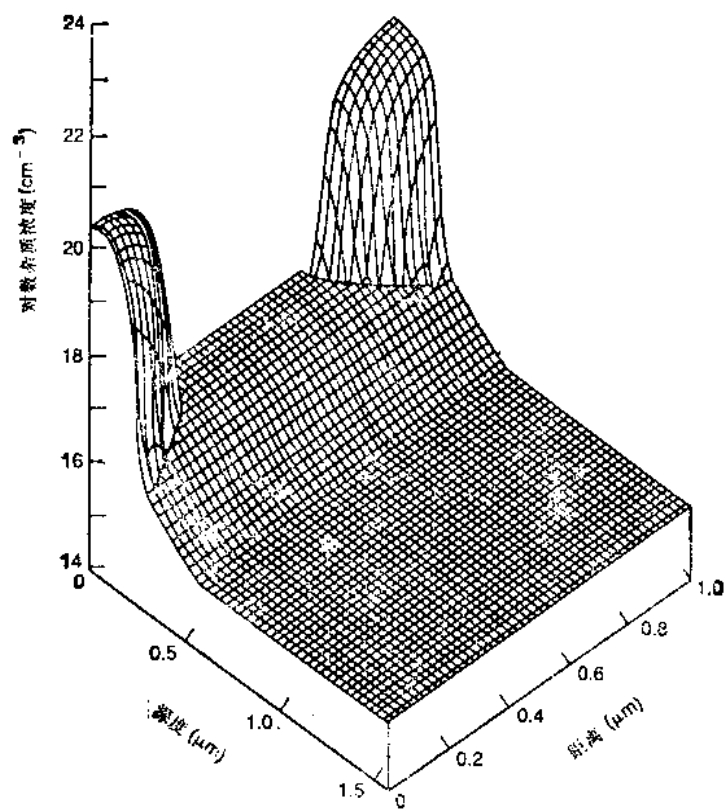


图2.55 1 μ m增强型晶体管模拟的杂质浓度分布曲面图

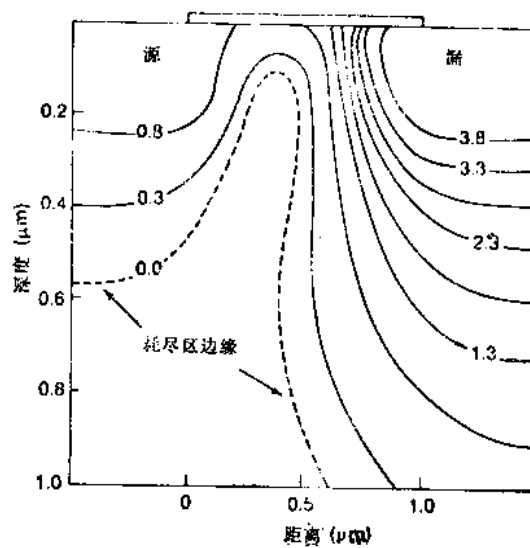


图2.56 图2.55中器件在饱和情况下的静电电位分布 ($V_{GS} = 1\text{ V}$, $V_{DS} = 3\text{ V}$, $V_{SS} = 0\text{ V}$, $T_{OX} = 250\text{ \AA}$, $V_{TX} = 0.6\text{ V}$)

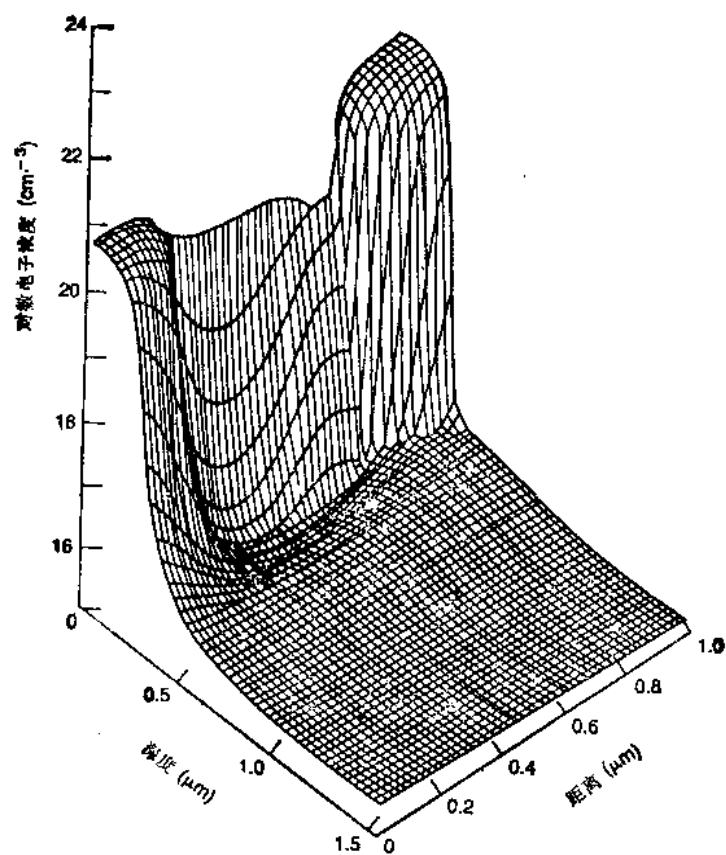


图2.57 与图2.56同样偏置条件下的电子分布

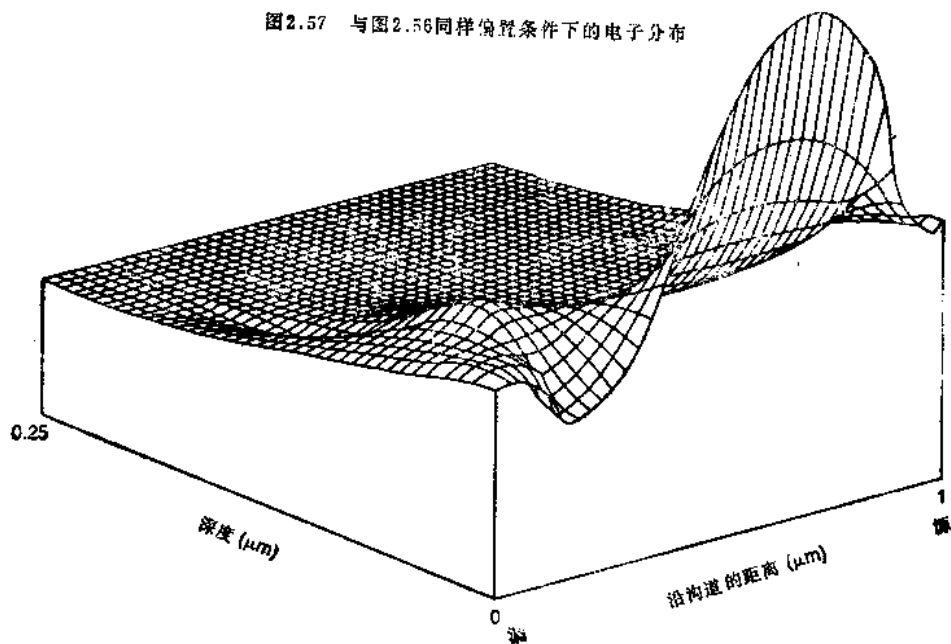


图2.58 栅下源漏方向电子流密度分布。在源附近电流主要在表面，但在漏附近电流渗入衬底。器件的偏置条件同图2.56

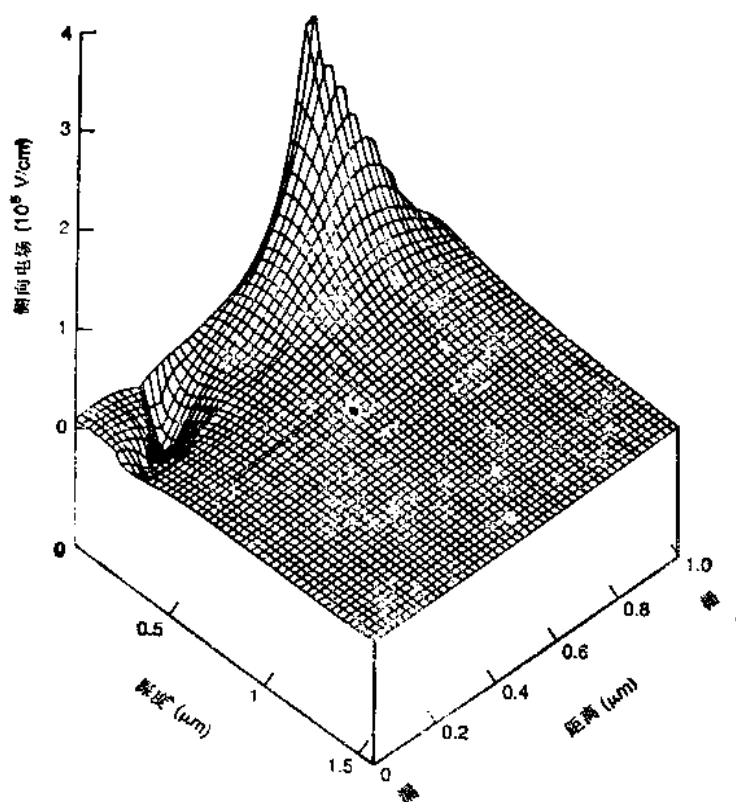


图2.59 在与图2.56—2.58同样的偏置条件下的侧向电场。电场的峰值位于漏附近

互连线远非理想情况。导体既有电阻也有电容。边缘电容可能与平行板电容大小差不多。长RC线遵守扩散方程，具有延迟随长度平方成正比的特征。

散热是一个大问题，特别是nMOS工艺更是如此。由于这一原因，芯片功耗被限制在1—5W以内。

半导体器件物理构成了VLSI性能的一组限制条件。其它限制包括制造工艺限制（第三章）及电学范畴与数字抽象之间的可靠变换（第四章）。

习 题

- 2.1 有一工艺， $T_{OX}=250\text{\AA}$ ，衬底电阻率为 $50\Omega\cdot\text{cm}$ 。请计算 ϕ_{FP} ， $2\Phi_{FP}$ ， X_{Pmax} ， V_{FB} ， V_{TE} ， δ 及 δ 。设 $N_f=1.5\times 10^{11}\text{cm}^{-2}$ ， $Q_{it}=\rho(z)=0$ 。在 $V_{BB}=0$ 及 -2.5V 的情况下分别进行计算。

这些计算适用于“自然”晶体管。现在我们要研究由多晶及金属线在衬底上形成的寄生晶体管的一些性质。假定金属为铝，多晶为简并掺杂的n型，栅氧化层为场氧化层及CVD氧化层。请计算 V_{FB} ， γ 及 V_{TX} （设 $V_{BB}=0$ ， $T_{CVD}=5000\text{\AA}$ ， $T_{BK}=7000\text{\AA}$ ）

如果你的计算正确，你将发现这些晶体管大概是永远导通的。为了解决这一问题，在场氧化层下加一次场注入以提高这些器件的阈值。这也提高了与这些器件有关的衬底杂质浓度。为了将场开启电压提高到15V以上所需的 N_s 值是多少？在这一杂质浓度下， γ 和雪崩击穿电压 V_{BR} 是多少？ Δ

2.2 图 2.60 所示三管存储单元, 可以利用体效应进行工作, 解释如何从这一存储单元读出和写入。△

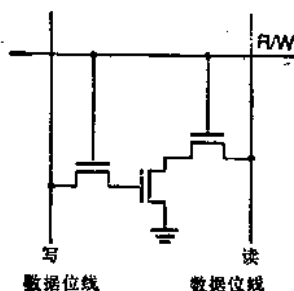


图 2.60

2.3 在第 2.2 节中对 Q_B 中的体电荷项作了简化, 但通过积分得到 Q_B 的完整方程是可能的。

请通过 Q_B 的完整方程推出

$$I_{DS} = SK' \{ (V_G - V_{FB} - 2\Phi_{FP} - V_S)^2 - (V_G - V_{FB} - 2\Phi_{FP} - V_D)^2 - \frac{4\gamma}{3} [(V_D + 2\Phi_{FP})^{1.5} - (V_S + 2\Phi_{FP})^{1.5}] \}$$

为找出饱和点 V_{Dsat} , 令 $Q_N = 0$, 推出

$$V_{Dsat} = V_G - V_{FB} - 2\Phi_{FP} + \frac{1}{2}\gamma^2 [1 - \sqrt{1 + 4(V_G - V_{FB})/\gamma}]$$

2.4 按我们的理论, 当 $V_G > V_{Tx}$ 时形成沟道。但若 V_S 和 V_D 大于零, 只要 $V_G < V_S + V_{Tx}$ 就不会有电流。那么, 如果存在沟道, 为什么会没有电流?

2.5 定义 $g_m \equiv \partial i_D / \partial v_{GS}$, $g_{mb} \equiv \partial i_D / \partial v_{BS}$, 证明 $g_{mb} = b g_m$ 。

2.6 热导所遵守的方程与电容的方程是完全一样的。对于薄的导线 ($W/H \ll 10$), 由于热的扩展, 热阻实际上比方程 (2.175) 所给出的要小。在第 2.4 节中我们看到由于电力线的扩展, 导线电容增加。利用这些结果推导一个 ΔT 的较精确的表达式。

2.7 图 2.6 画出了一个 MOS 电容的 $C-V$ 特性。这一电容可分为三个组成部分, 如图 2.29 所示。利用方程 (2.155) 和 (2.156) 所给出的 C_N 和 C_B 的定义, 画出这三个电容随 V_G 的变化。

2.8 通过建立合理的 SPICE 模型卡片 (设 $\gamma = 0.6$, 也许可以用附录给出的模型来编辑), 从一个方面来比较 SPICE level 2 和 level 3 模型。把一个 1 米宽 3 米长的晶体管 (我们可以肯定这一晶体管没有短沟效应) 模型也放在一起, 并且看 $V_{GS} = V_{DS} = 5V$ 下的电流。然后把晶体管分解为三个 1 米 $\times$ 1 米的互相串联的晶体管, 它们的栅连在一起, 观察在同一偏置条件下的电流。用 level 2 和 level 3 两个 SPICE 模型来进行模拟。解释你得到的 (有些奇怪的) 结果。△

2.9 为什么 MOS 电容值不等于 $V_G = V_{FB}$ 时的 C_{ox} 值 (启示: 这在课文中未讨论过。此问题与耗尽近似以及一个称之为德拜长度的量有关)。

2.10 整理例 2.9 中的 ROM 模型, 观察在设计中作下列不同改变时其效果的百分比值 (每次改变后仍回到原来的模型):

- 令 $R_{word} = 0$,
- 令 $C_c = 0$,
- 每 7 位放一条接地带,
- 从正中驱动 ROM。△

2.11 硅中一般施主与受主原子的电离电位在 40mV 的量级而室温约对应于 25mV, 但室温下几乎所有施主与受主都是离化的, 为什么? △△

2.12 利用图 2.62 所示模型来模拟图 2.61 所示晶体管需要多少段? 要求的精度为 5%。请采用附录中的 $2\mu m$ nMOS 增强型晶体管模型。

2.13 在一个改进的 MOS 工艺中, 我们试着使 γ 好一些, 哪些方面会变坏?

2.14 Charlie Analog 宣称当 $T_{ox} = 100 \text{ \AA}$ 而不是 $1000 \text{ \AA}$ 时他可大大改进对 MOS 晶体管阈值的控制, 这是否合理? 为什么?

2.15 Janet DeVice 宣称在 $2\mu m$ 工艺的例子中所采用的 θ 值太低了。用该章中的方程估算这些工艺的 θ 值。问: 她对吗? 当 $V_{GS} = V_{Tx}$, 氧化层中电场的表达式是怎样的?

2.16 电场 $\mathcal{E}$ 代表作用在电子上的力, 为什么它引起一定的速度 ($v = \mu_e \mathcal{E}$) 而不是一定的加速度?

2.17 假定我们采用 $2\mu m$ CMOS 工艺并把所有尺寸 (包括耗尽区宽度) 减为一半, 此外一切不变,

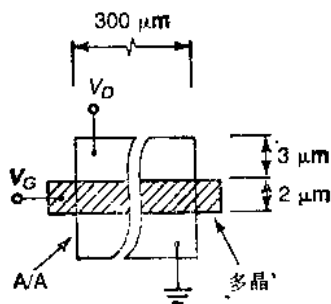


图 2.61

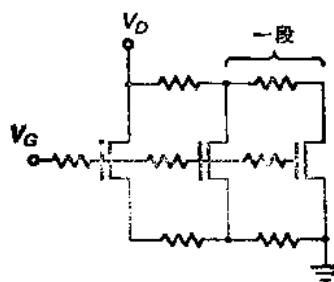


图 2.62

并使之仍工作在 5V。假定器件不击穿或穿通，设器件形状保持不变，问一个长度为最短而宽度不短的器件缩小前后其不同工作区的电流比较结果如何？例如，对 $V_{GS} = 5V$ 的 n 沟器件，在 $V_{DS} = 0.1V$ 和 $V_{DS} = 5V$ 时分别比较 $2\mu m$ 工艺中的 $20/2$ 器件与缩小的 $1\mu m$ 工艺中的 $10/1$ 器件。当 $V_{GS} = 0.5V$ 时情况如何？并就答案进行解释哪些是重要的效应？

参 考 文 献

- [Akers 82] L. A. Akers and J. J. Sanchez, "Threshold Voltage Models of Short, Narrow, and Small Geometry MOSFETs, A Review," *Solid-State Electronics* 25, 621—641, 1982.
- [Baccarani 84] G. Baccarani, M. Wordeman, and R. H. Dennard, "Generalized Scaling Theory and Its Application to a $1\frac{1}{2}$ Micrometer MOSFET Design," *IEEE Trans. Electron Devices* ED-31: 452—462, 1984.
- [Bain 82] L. I. Bain, private communication, 1982.
- [Balk 74] P. Balk, "Layered Dielectrics in the MOS Technology," *Inst. Phys. Conf. Series* 1973 19, 51—82, 1974.
- [Barnes 79] J. J. Barnes, K. Shimohigashi, and R. W. Dutton, "Short-Channel MOSFET's in the Punchthrough Current Mode," *IEEE J. Solid-State Circuits* SC-14, 368—374, 1979.
- [Baum 70] G. Baum and H. Beneking, "Drift Velocity Saturation in MOS Transistors," *IEEE Trans. Electron Devices* ED-17, 481—482, 1970.
- [Bell 70] *Physical Design of Electronic Systems* 1, "Design Technology, Pt. II, Thermal Design," Members of Technical Staff, Bell Telephone Laboratories, Prentice-Hall, Englewood Cliffs, N.J., 1970.
- [Black 69] J. R. Black, "Electromigration Failure Modes in Aluminum Metalization for Semiconductor Devices," *Proc. IEEE* 57, 1587—1594, 1969.
- [Botchek 83] C. M. Botchek, *VLSI Basic MOS Engineering* 1, Pacific Technology Group, Saratoga Calif., 1983.
- [Brews 80] J. R. Brews, W. Fichtner, E. H. Nicollian, and S. M. Sze, "Generalized Guide for MOSFET Miniaturization," *IEEE Trans. Electron Devices Lett.* EDL-1, 2—4, 1980.
- [Chua 81] S. J. Chua, "Current and Resistivity Dependence of Electromigration from a Statistical Analysis of Metalization Failure Data," *Solid State Electronics* 24, 173—178, 1981.
- [Colinge 81] J. P. Colinge, E. Demoulin, P. Delannay, and M. Lobet, "Grain Size and Resistivity of LPCVD Polycrystalline Silicon Films," *J. Electrochem. Soc., Solid-State Science and Technology* 128, 2009—2014, 1981.
- [Dang 79] L. M. Dang, "A Simple Current Model for Short-Channel IGFET and Its Application to Circuit Simulation," *IEEE J. Solid-State Circuits* SC-14, 358—367, 1979.
- [Dang 81] R. L. M. Dang and N. Shigyo, "Coupling Capacitances for Two-Dimensional Wires," *IEEE Trans. Electron Devices Lett.* EDL-2, 196—197, 1981.

- [Davis 81] J. R. Davis, *Instabilities in MOS Devices*, Gordon and Breach, New York, 1981.
- [Dennard 74] R. H. Dennard, F. H. Gaensslen, H. N. Yu, V. L. Rideout, E. Bassons, and A. R. LeBlanc, "Design of Ion-Implanted MOSFET's with Very Small Dimensions," *IEEE J. Solid-State Circuits* SC-9, 256—267, 1974.
- [Elmasry 82] M. I. Elmasry, "Capacitance Calculations in MOSFET VLSI," *IEEE Trans. Electron Devices Lett*, EDL-3, 6—7, Jan, 1982.
- [Estreich 80] D. B. Estreich, "The Physics and Modeling of Latch-up and CMOS Integrated Circuits," Stanford Electronics Labs, Stanford, Calif., Tech. Rep. G201-9, November 1980.
- [Estreich 82] D. B. Estreich and R. W. Dutton, "Modeling Latch-up in CMOS Integrated Circuits," *IEEE Trans. Computer-Aided Design CAD-1*, 157—162, 1982.
- [Fang 70] F. F. Fang and A. B. Fowler, "Hot Electron Effects and Saturation Velocities in Silicon Inversion Layers," *J. Appl. Phys.* 44, 1825—1831, 1970.
- [Gibbons 75] J. F. Gibbons, W. S. Johnson, and S. W. Mylroie, *Projected Range Statistics, Semiconductors and Related Materials*. (2nd ed.), Halstead Press, Stroudsburg, Pa., 1975.
- [Gibbons 80] J. F. Gibbons and K. F. Lee, "One-Gate-Wide CMOS Inverter on Laser-Recrystallized Polysilicon," *IEEE Trans. Electron Devices Lett*, EDL-1, 117—118, 1980.
- [Glaser 77] A. B. Glaser and G. E. Subak-Sharpe, *Integrated Circuit Engineering*, Addison-Wesley, Reading, Mass., 1977.
- [Goetzberger 66] A. Goetzberger, "Ideal MOS Curves for Silicon," *Bell Syst. Tech. J.* 45, 1097—1123, 1966.
- [Gray 69] P. E. Gray and C. L. Scarle, *Electronic Principles, Physics, Models, and Circuits*, Wiley, New York, 1969.
- [Greeneich 83] E. W. Greeneich, "An Analytical Model for the Gate Capacitance of Small-Geometry MOS Structures," *IEEE Trans. Electron Devices* ED-30, 1838—1839, 1983.
- [Gregory 73] B. L. Gregory and B. D. Shafer, "Latch-up in CMOS Integrated Circuits," *IEEE Trans. on Nuclear Science* NS-20, 293—299, 1973.
- [Grove 67] A. S. Grove, *Physics and Technology of Semiconductor Devices*, Wiley, New York, 1967.
- [Hall 68] P. M. Hall, "Resistance Calculations for Thin Film Patterns," *Thin Solid Films, An International Journal on their Science and Technology* 1, 277—285, Elsevier, Amsterdam, 1967/68.
- [Hamilton 75] D. J. Hamilton and W. G. Howard, *Basic Integrated Circuit Engineering*, McGraw-Hill, 1975.
- [Heil 35] O. Heil, British Patent 439, 457 (1935).
- [Heurle71] F. M. D. Henle. "Electromigration and Failure in Electronics, An Introduction," *Proc. IEEE* 59, 1409—1418, 1971.
- [Hoeneisen 72] B. Hoeneisen and C. A. Mead, "Fundamental Limitations in Microelectronics I. MOS Technology," *Solid-State Electron.* 15, 819—829, 1972.
- [Horowitz 83] M. Horowitz and R. W. Dutton, "Resistance Extraction from Mask Layout Data," *IEEE Trans. Computer-Aided Design CAD-2*, 145—150, 1983.
- [Irvin 62] J. C. Irvin, "Resistivity of Bulk Silicon and of Diffused Layers in Silicon," *Bell Sys. Tech. J.* 41, 388, 1962.
- [Jacoboni 77] C. Jacoboni, C. Canali, G. Ottaviani, and A. A. Quaranta, "A Review of Some Charge Transport Properties in Silicon," *Solid State Electronics* 20, 77—89, 1977.
- [Keyes 75] R. W. Keyes, "Physical Limits in Digital Electronics," *Proc. IEEE* 63, 740—767, 1975.
- [Kittel 76] C. Kittel, *Introduction to Solid State Physics*, p. 171, Wiley, New York, 1976.
- [Kriegler 74] R. J. Kriegler, "Ion Instabilities in MOS Structures," *Proc. 12th IEEE Reliability Phys. Sym.*, pp. 250—259, New York, 1974.
- [Kroell 76] K. E. Kroell and G. H. Ackermann, "Threshold Voltage of Narrow Channel Field Effect Transistors," *Solid-State Electron.* 19, 77—81, 1976.
- [Lewis 84] E. T. Lewis, "An Analysis of Interconnect Line Capacitance and Coupling for VLSI Circuits," *Solid State Electronics*, to be published.

- [Lilienfeld 80] J. E. Lilienfeld, U.S. Patent 1, 745, 175(1930).
- [Lu 81] N. C.-C. Lu, L. Gerzberg, C.-Y. Lu, and J. D. Meindl, "Modeling and Optimization of Monolithic Polycrystalline Silicon Resistors," *IEEE Trans. Electron Devices* ED-28: 818-830, 1981.
- [Masuda 79] H. Masuda, M. Nakai, and M. Kubo, "Characteristics and Limitations of Scaled Down MOSFET's Due to Two Dimensional Field Effect," *IEEE Trans. Electron Devices* ED-26: 980-986, 1979.
- [Merkel 77] G. Merkel, "CAD Models of MOSFETs," *Process and Device Modeling for Integrated Circuit Design*, Wiele, Engl, and Jaspers (eds.), Noordhoff, Leyden, the Netherlands, 1977.
- [Meyer 71] J. E. Meyer, "MOS Models and Circuit Simulation," *RCA Rev* 32: 42-63, March 1971.
- [Muller 77] R. S. Muller and T. I. Kamins, *Device Electronics for Integrated Circuits*, Wiley, New York, 1977.
- [Murtuza 82] M. Murtuza, "Computer-generated Models Abridge Thermal Analysis of Packaged VLSI," *Electronics* 55: 145-148, Feb. 10, 1982.
- [Nicollian 82] E. H. Nicollian and J. R. Brews, *MOS Physics and Technology*, Wiley, New York, 1982.
- [Ochoa 79] A. Ochoa, W. Dawes, and D. Estreich, "Latchup Control in CMOS Integrated Circuits," *IEEE Trans. Nucl. Sci.* NS-26: 5065-5068, 1979.
- [Ogura 82] S. Ogura, C. F. Codella, N. Rovedo, J. F. Shepard, and J. Riseman, "A Half Micron MOSFET Using Double Implanted LDD," *IEEE International Electron Device Meeting*, pp. 718-721, San Francisco, 1982.
- [Paulos 82] J. J. Paulos, D. A. Antoniadis, and Y. P. Tsividis, "Measurement of Intrinsic Capacitances of MOS Transistors," *IEEE J. Solid-State Circuits*, pp. 238-239, 1982.
- [Paulos 83] J. J. Paulos and D. A. Antoniadis, "Limitations of Quasi-Static Capacitance Models for the MOS Transistor," *IEEE Trans. Electron Devices Lett.* EDL-4: 221-224, 1983.
- [Protonotarios 67] E. N. Protonotarios and O. Wing, "Theory of Nonuniform RC Lines. Part II, Analytic Properties in the Time Domain," *IEEE Trans. Circuit Theory* CT-14: 13-20, 1967.
- [Ratnakumar 81] K. N. Ratnakumar, J. D. Meindl, and D. L. Scharfetter, "New IGFET Short-Channel Threshold Voltage Model," *IEEE International Electron Device Meeting*, pp. 204-206, Washington, D. C., 1981.
- [Rubinstein 83] J. Rubinstein, P. Penfield, Jr., and M. A. Horowitz, "Signal Delay in RC Tree Networks," *IEEE Trans. Computer-Aided Design* CAD-2: pp. 202-211, 1983.
- [Rung 82] R. D. Rung, H. Momose, and Y. Nagakubo, "Deep Trench Isolated CMOS Devices," *IEEE International Electron Device Meeting*, pp. 237-240, San Francisco, 1982.
- [Sabnis 79] A. G. Sabnis and J. T. Clemens, "Characterization of Electron Mobility in the Inverted <100> Si Surface," *IEEE International Electron Device Meeting*, pp. 18-21, Washington, D. C., 1979.
- [Sah 64] C. T. Sah, "Characteristics of the Metal-Oxide-Semiconductor Transistors," *IEEE Trans. Electron Devices* ED-11: 324-345, 1964.
- [Schroeder 80] J. E. Schroeder and A. Ochoa, Jr., and P. V. Dressendorfer, "Latch-up Elimination in Bulk CMOS LSI Circuits," *IEEE Trans. on Nuclear Science* NS-27: 1735-1738, 1980.
- [Schwarz 83] S. A. Schwarz and S. E. Russek, "Semi-Empirical Equations for Electron Velocity in Silicon, Part II MOS Inversion Layer," *IEEE Trans. Electron Devices* ED-30: 1634-1639, 1983.
- [Serhan 83] G. I. Serhan and S.-Y. Yu, "A Simple Charge-Based Model for MOS Transistor Capacitances: A New Production Tool," *IEEE Trans. Computer-Aided Design* CAD-2: 48-51, January 1983.
- [Seto 75] J. Y. W. Seto, "The Electrical Properties of Polycrystalline Silicon Films," *J. Applied Phys.* 46: 5247-5254, 1975.
- [Shickman 68] H. Shickman and D. Hodges, "Modeling and Simulation of Insulated-Gate Field-Effect Transistor Switching Circuits," *IEEE J. Solid State Circuits* SC-3: 285-289, 1968.
- [Shockley 48] W. Shockley and G. L. Pearson, "Modulation of Conductance of Thin Films of Semiconductors by Surface Charges," *Phys. Rev.* 74: 232, 1948.

- [Singh 76] B. R. Singh and K. Singh, "Instability Phenomena in Thin Insulating Films on Silicon," *Microelectronics and Reliability* 15, 385—398, 1976.
- [Smith 80] P. Smith, M. Inoue, and J. Frey, "Electron Velocity in Si and GaAs at Very High Electric Fields," *Appl. Phys. Lett.* 37, 797—798, 1980.
- [Snow 68] E. H. Snow and B. E. Deal, "Polarization Effects in Insulating Films on Silicon: A Review," *Trans. Met. Soc. AIME* 242, 512—522, 1968.
- [Sodini 82] C. G. Sodini, T. W. Ekstedt, and J. L. Moll, "Charge Accumulation and Mobility in Thin Dielectric MOS Transistors," *Solid-State Electronics* 25, 833—841, 1982.
- [Swanson 72] R. M. Swanson and J. D. Meindl, "Ion-Implanted Complementary MOS Transistors in Low-Voltage Circuits," *IEEE J. Solid-State Circuits* SC-7, 146—153, 1972.
- [Sze 81] S. M. Sze, *Physics of Semiconductor Devices*, Wiley, New York, 1981.
- [Troutman 79] R. R. Troutman, "VLSI Limitations from Drain-Induced Barrier Lowering," *IEEE J. Solid-State Circuits* SC-14, 383—390, 1979. (Also punch-through)
- [Troutman 83] R. R. Troutman, "Epitaxial Layer Enhancement of n-Well Guard Rings for CMOS Circuits," *IEEE Trans. Electron Devices* EDL-4, 438—440, 1983.
- [Troutman 84] R. R. Troutman and H. P. Zappe, "Layout and Bias Considerations for Preventing Transiently Triggered Latchup in CMOS," *IEEE Trans. Electron Devices* ED-31, 315—321, 1984.
- [Tsividis 80] Y. Tsividis, "Relation Between Incremental Intrinsic Capacitances and Transconductances in MOS Transistors," *IEEE Trans. Electron Devices* ED-27, 946—948, 1980.
- [Tucherman 82] D. B. Tucherman and R. F. W. Pease, "High-Performance Heat Sinking for VLSI," *IEEE Trans. Electron Devices* EDL-2, 126—129, 1981.
- [Vladimirescu 80] A. Vladimirescu and S. Liu, "The Simulation of MOS Integrated Circuits using SPI-CE2," Memo. UCB/ERL M80/7, University of California, Berkeley, October, 1980.
- [Walczyk 83] F. Walczyk and J. Rubenstein, "A Merged CMOS/Bipolar VLSI Process," *IEEE International Electron Device Meeting*, pp. 59—62, Washington, D. C., 1983.
- [Wallmark 74] J. T. Wallmark and L. G. Carlstedt, *Field-Effect Transistors in Integrated Circuits*, Halsted Press, New York, 1974.
- [Ward 78] D. E. Ward and R. W. Dutton, "A Charge-Oriented Model for MOS Transistor Capacitances," *IEEE J. Solid-State Circuits* SC-13, 703—707, 1978.
- [Ward 82] D. E. Ward and K. Doganis, "Optimized Extraction of MOS Model Parameters," *IEEE Trans. Computer-Aided Design* CAD-1, 163—168, 1982.
- [Woods 74] M. H. Woods, "Instabilities in Double Dielectric Structures," *Proc. 12th IEEE Reliability Phys. Sym.*, pp. 259—266, New York, 1974.
- [Wordeman 79] M. R. Wordeman, "Characterization of Depletion-mode MOSFETs," *IEEE International Electron Device Meeting*, pp. 26—29, Washington, D. C., 1979.
- [Yang 78] E. S. Yang, *Fundamentals of Semiconductor Devices*, McGraw-Hill, New York, 1978.
- [Yang 82] P. Yang and P. K. Chatterjee, "SPICE Modeling for Small Geometry MOSFET Circuits," *IEEE Trans. Computer-Aided Design* CAD-1, 169—182, 1982.
- [Yan 74] L. D. Yan, "A Simple Theory to Predict the Threshold Voltage of Short-channel IGFE-T's," *Solid State Electronics* 17, 1059—1063, 1974.
- [Yuan 82] C. P. Yuan and T. N. Trick, "A Simple Formula for the Estimation of the Capacitance of Two-Dimensional Interconnects in VLSI Circuits," *IEEE Trans. Electron Devices* EDL-3, 391—393, 1982.

第三章 芯片工艺

建立一个 VLSI 的制造设施,可能要耗资几千万至几万万美元,用这些设施和日益复杂的加工技术及设备,生产出来的一个元件售价约为 5 美元。在这一章中,我们将讨论与 VLSI 芯片制造及 VLSI 芯片整体性有关的问题。大批量生产是工艺技术的主要宗旨。从历史上看,由生产工艺方面采取措施提高经济效益,往往都把注意力放在(在产品满足所要求的规范条件下)优化加工工艺、采用更大直径的硅片、芯片面积最小化以及使用最少的掩模层次等方面。过分强调上述后两项措施将会加长设计时间。而这一点也是与成本相关的。怎样使产品的总成本最低往往是很难判断的。

加工一个 VLSI,要用一百多步独立的加工工序。从某种意义上说,这些工序是紧密相关的,改变任何一步的工艺,都会使从表面上看来似乎无关的工艺发生变化。例如,氧化生长时间会影响杂质分布,而反过来氧化生长速率却又依赖于杂质浓度。这种现象使得难以孤立地只改变一个参数而不影响其它参数。出于经济方面的考虑,大多数工艺师把目标放在产品成本的最优化上,而不是单纯处理各工序间的独立性问题。¹⁾

在这一章中,我们将介绍加工工艺和 VLSI 性能之间的联系。对于制造 VLSI 来说,不仅工艺技术,而且加工过程中的不定性和公差都是必须先决条件。单靠生产工艺是不可能准确无误地制造出完全一致的产品。

与艺术家的创作一样,电路设计师对于有关问题了解得越充分,他的设计结果就会越好。在本章中,将讨论一些比较重要的 MOS 加工工艺技术。

3.1 半导体器件的制造技术

在这一节中,我们要分析一系列制造工艺技术,工艺工程师可用这些技术来生产 VLSI 芯片 [Sze 83, Colclaser 80, Giaser 77]。在下一节则分析怎样把这些技术组合起来制做 MOS 集成电路。

晶体生长

单晶硅锭生长是衬底准备工艺的第一步。通常采用 Czochralski [Rea 81] 法,从充满熔融硅和合适的掺杂剂的坩埚中拉取硅单晶。坩埚用石墨做成。单晶生长要在惰性气氛(如氩气)中进行,用高频或电阻加热器提供能量。单晶径向电阻率的变化大约为 $\pm 10\%$ 。用这种技术可以得到电阻率从 $30\text{k}\Omega\cdot\text{cm}$ 到 $0.0001\Omega\cdot\text{cm}$ 的硅单晶,为了得到更高电阻率的硅单晶则要配以区域熔化提纯技术。

1) 由于整体优化失掉了灵活性,因而近期在设计工艺“模块”方面进行了一些研究。针对不同的设计项目,可以用不同的方式把这些“模块”组合起来。

然后，把单晶磨制成标准圆柱体，用X射线衍射确定晶向¹⁾并在圆柱体上磨一个平台以标记这个晶向。图3.1示出了P型<100>单晶的平台。然后把单晶硅锭锯成片状，再通过腐蚀去掉划伤的表面并把片子背面打毛，正面则抛光至粗糙度小于 $\pm 6\mu\text{m}$ ，最后清洗硅片，硅片的最终厚度约为 $300\mu\text{m}$ 。

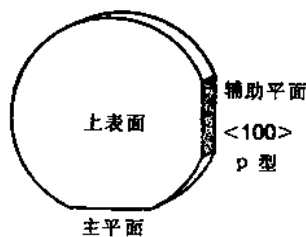


图3.1 主平面和标志P型<100>硅片晶向的辅助平面。抛光的面朝上

在需要的情况下，接下来可以生长一层外延层²⁾。在外延工艺中，使某种气体(例如 SiH_4)通过硅片表面，并使 SiH_4 分解。利用这种工艺技术，可在单晶片上生长厚度为 0.5 至 $100\mu\text{m}$ 的单晶硅膜。外延层的生长速率与温度有关。在一种较为成熟的工艺中， 1200°C 分解 SiCl_4 时，其生长速率的量级为 $1\mu\text{m}/\text{min}$ 。在 1080°C 左右用 SiH_2Cl_2 分解时，其生长速率约为前者的 $1/3$ 。用 SiH_4 可在 1000°C 左右生长外延层，降低使用的温度，以便使杂质的迁移尽可能地小些。可以从简并掺杂的n型衬底上生长出电阻率为 $50\Omega\cdot\text{cm}$ 的外延层，其过渡区约为 $2\mu\text{m}$ 。

外延层的重要优点之一是它可以具有与衬底非常不同的掺杂浓度。硼是外延层中典型的n型杂质，自然，硼是典型的P型杂质。在外延生长之前，可以向衬底扩散一个埋层，硼和锑是典型的埋层掺杂用杂质。这种工艺技术在双极型电路中非常普遍，在制作CMOS电路时有时也采用，例如，所谓“反转阱”(retrograde well)，其中掺杂浓度从表面向内随距离增加而上升，这种结构有利于抑制自锁(Latch-up)效应。随着埋层掺杂的加重，可能会引起晶体损伤。由于外延是在高温下进行的，衬底中杂质的蒸气压就可能成为一个问题。处在气流上游的硅片会逸出杂质，而在下游的则吸收这些杂质。此外，气流中杂质也会耗尽。有时，在片子背面生长一层氧化层，用来限制杂质逸出到气体中去。考虑到这些效应，在外延生长中，必须不断地调整气流和硅片的温度。在双阱CMOS和其它抑制自锁效应的CMOS工艺中，外延生长是关键工序。

光复印和掩模制作

光刻工艺把许多使某种材料有选择地从硅片上除去或加到硅片上的工艺技术联系起来。掩模包含着要转移到硅片上去的那些信息。几何尺寸约大于 $1\mu\text{m}$ 时，可以采用光学(实际上是紫外线)技术。最廉价的掩模是在对紫外光透明的玻璃上涂覆 $2\text{—}4\mu\text{m}$ 厚的照相感光乳胶制成，这些掩模往往用于接触式复印，这时，掩模和硅片紧密接触，因而掩模的磨损严重。接触式复印的优点是透镜和光学系统相当简单。

硬掩模是在玻璃上涂覆厚度为 $1000\text{—}2000\text{\AA}$ 的铬或氧化铁，这些掩模往往用在投影对准系统之中³⁾，氧化铁的优点在于它对可见光是半透明的而对紫外光却不透明，这就使得掩模对准更为容易，X射线所用的掩模常常是在铍或硅衬底上的金层中腐蚀出图

1) 硅与金刚石具有同样的晶体结构。如金刚石一样，可以沿许多不同的平面切开硅单晶，但并不是所有这些平面都是一样的。例如，对于不同晶向表面处的原子密度是不同的。两个不同的晶向用<111>和<100>标示。

2) 外延层是生长在衬底表面上的一薄层材料。这个字来自字根“epi”，其含义是“在上边”。

3) 在投影对准系统中，掩模上的图象是被投影到片子上的，片子与掩模从不接触。

案。由于X射线波长较短，所以从紫外曝光发展到X射线曝光，可以得到更高的分辨率。

掩模图案可以用光学方法或用电子束制版机产生。光学方法是使用一个精密的机械装置，通过一个可调光阑把一些矩形曝光到照相底版上，一旦芯片的一层掩模曝光后，再通过步进重复技术将它多次重复转移，在一个掩模内拷贝了成百个管芯，这就是70年代的标准技术。用做好的掩模在硅片上把所有的芯片同时印出来。对于硅片上的所有芯片，许多工艺步骤都是同时实现的，所以工艺的进步都倾向于使用较大的硅片以提高生产效率。

除了掩模图案采用电子束直接绘出之外，电子束技术和上述光学技术是很相似的。电子束是由计算机驱动的，与光学或X射线技术不同，它必须连续地绘出每一个像素，然后就用步进重复技术，或用电子束制版机在掩模上所有地方绘出图像。由电子束制版机作出的掩模(所谓母版)是非常昂贵的，但比起用光学方法制作的掩模质量较好。图3.2示出了一个掩模。当直接使用掩模原版时，要采用投影对准器。由于掩模不接触芯片，所以使用期一定会更长。投影对准器的透镜是非常复杂的，且随着硅片直径的加大而更为复杂。因为曝光是通过一束窄条紫外光移过硅片实现的，所以早期制作的投影对准器具有方向性，这就是说，在 z 方向上的几何图形比 y 方向的图形更为精确完好。然而这个特点没有继承下来。在用电子束描绘得到的掩模上，并没有限制每一个芯片位置必须含有同样的信息，所以可以用电子束技术制作含有多种设计图形的掩模 [Conway 80,

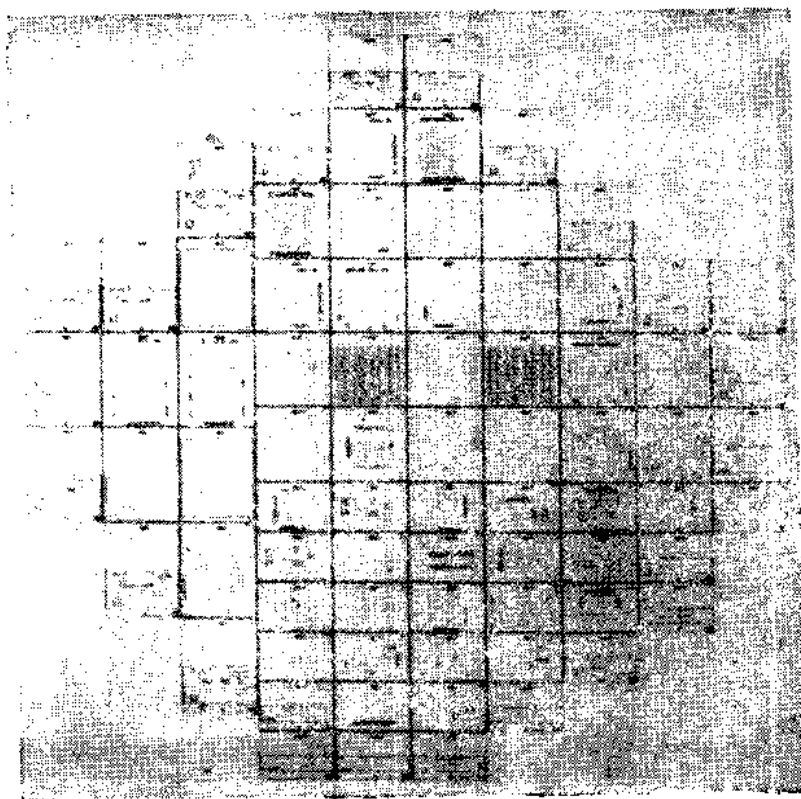


图3.2 4 μm nMOS技术的有源区掩模照片 (经MOSIS许可使用)

Jansen 81]。亦即,在一个掩模上允许有许多不同的芯片设计图形,这是硅芯片加工厂发展的关键。在这些工厂中,可以用同一套掩模,在同一硅片上同时加工许多不同用户所设计的电路。电子束制造的掩模(其几何尺寸通常是硅片上最终尺寸的5或10倍)用另一种技术可以直接在硅片上而不是在第二个掩模上步进重复,这种类型的对准器被称作硅片步进器。硅片步进器允许在每个芯片位置上进行对准,由于曝光范围较小,所以可以使用更好的镜头¹⁾。还应当指出,电子束机可以直接在硅片上绘出图形,而根本不用任何掩模。这种技术当然要慢一些,但是在某些研究领域却很有用。

先在硅片上涂覆一层光致抗蚀剂,然后通过掩模使光致抗蚀剂曝光,图形就从掩模转移到了硅片上。光致抗蚀剂是一种有机聚合物,它对于光或电子束是很敏感的。光致抗蚀剂通常有两种类型:正性和负性。在负性光致抗蚀剂中,凡紫外线照射到的部位,就使聚合物的键交联,并形成长链有机分子。对正性光致抗蚀剂则相反,只要受到光线照射的部位,材料分子的链就被断开,使之成为更小的分子。不管哪种抗蚀剂,都能通过在弱有机溶液中漂洗的方法进行显影,这种有机溶液只能溶解较短小的分子,这样一来,对于正性光致抗蚀剂,曝光部位被去掉。而负性抗蚀剂的情况则正好相反。由于光线的反射作用,用正性光致抗蚀剂时,线条会变细,而用负性光致抗蚀剂时,线条则会加宽。为此,要获得精细几何尺寸时,应优先选用正性光致抗蚀剂。

某些特殊的光致抗蚀剂对X射线或电子束比较敏感,而对紫外线却不太敏感。有些抗蚀剂则是无机材料。抗蚀剂的主要质量标准包括它的粘附力、分辨率、灵敏度和对溶液的抗蚀能力。在一种单一的抗蚀剂中,往往难以达到所需要的全部要求。因此,有时采用多层抗蚀剂系统。采用这些系统时,只使顶层抗蚀剂曝光成掩模的图案,再用这层抗蚀剂作为掩模。经化学腐蚀得到下一层图形。由于底层抗蚀剂会填满硅片表面的那些凹坑,因而顶层抗蚀剂就处在一个十分平滑的表面上。因此多层抗蚀剂比单层抗蚀剂的分辨率大约高出50%。在200℃以上,不能使用有机光致抗蚀剂。

把一滴光致抗蚀剂滴到置于匀胶器上的硅片上,控制速度和加速度使硅片旋转,光致抗蚀剂就分散成为近乎均匀的薄层。在掩模对准设备上把片子上的对准符号与掩模上的对准符号相互对准并进行曝光。对于许多工艺线,这一步往往都是瓶颈(细脖子环节),用这种方法可以进行层与层之间的定位。对手工操作的设备,操作人员通过立体显微镜观察掩模和芯片,通常用一些同心的方块形对准符号作为对准的标记。在更现代化的机器中,比如硅片步进器,在掩模上有计算机可以读出的对准符号,因而可使对准自动化。因为片子只能一片一片地曝光,所以掩模对准工艺是比其它工艺更为昂贵的工艺。比如外延生长工艺,一次可以加工20片,就比较廉价。

在抗蚀剂曝光和显影之后再进行烘烤,以增加其粘附力并成为坚实的膜。这样的抗蚀剂就可以用来保护硅片表面,使它不受腐蚀剂的腐蚀。腐蚀后再把硅片放到能够破坏长聚合链的溶液中去掉抗蚀剂。

材料的选择性去除

在许多去除技术中,腐蚀法或许是最成熟的。化学法或所谓湿法腐蚀的好处是有优良的选择性。选择性好就是指腐蚀剂对所要去除材料的腐蚀性很强,而对不希望去除的

1) 在加工过程中,硅片可能会膨胀或收缩几个微米,这是采用较小曝光区的另一个原因。

材料则恰恰相反。例如，缓冲的 HF 能腐蚀 SiO_2 而不腐蚀 Si。用稀释的 HF 来剥除玻璃和铝上的 SiO_2 ，就可以得到如图 1.2(b) 中所示 T11 的显微照片。使用 HF 时必须极端小心，HF 不会烧伤皮肤但却会渗到骨头中去，杀死细胞组织和神经 [Straub 83]。显然，不能把 HF 贮存在玻璃瓶子里。

许多腐蚀剂能优先腐蚀一定晶面的硅。这就可以得到某些有意义的硅结构。例如，腐蚀技术是制造 VMOS 的基础。在某种常用的 CMOS 工艺中，用 KOH 腐蚀场区，如图 3.3 所示。KOH 可以很快地腐蚀 $\langle 100 \rangle$ 面的硅材料，却几乎不腐蚀 $\langle 111 \rangle$ 面硅材料。在生长氧化层时，表面几乎是平的，如图 3.4 所示。然而，“鸟嘴”仍然很严重。已有人研究过减小“鸟嘴”的方法 [Oldham 82, Chin 82]。

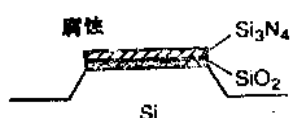


图 3.3 用 KOH 进行硅的优先腐蚀

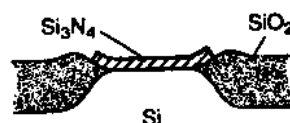


图 3.4 在场氧化中形成的“鸟嘴”。请注意是怎样借助于优先腐蚀而使顶部表面平整的。

等离子腐蚀、溅射和反应离子腐蚀的基本原理相同，都是在由高频或直流激活的气体中，利用受激离子去除硅表面上的某种材料。在溅射腐蚀中，气体是惰性的，材料是用机械方式去除的。等离子腐蚀的气体是化学激活的，材料的去除几乎是各向同性，就像湿法腐蚀一样。反应离子腐蚀（RIE）则是用化学激活离子进行溅射。RIE 的好处在于电场可以使离子垂直地撞击表面，腐蚀是各向异性的，从而可以得到极精细条宽所需的垂直陡峭边界。这些技术的主要缺点是选择性要比理想的差些。

在某些情况下，可以直接用光致抗蚀剂作为去除材料的掩模，而在另外的情况下，则需要采用间接的方法。例如，可以用光致抗蚀剂掩蔽来腐蚀 SiO_2 ，而 SiO_2 又可作为腐蚀 Si 的掩蔽膜。使用这种间接方法的一个重要实例是硅栅 CMOS 工艺中用来确定栅区，在多晶硅腐蚀后，多晶硅首先用作腐蚀栅氧化层的掩模，随后用来掩蔽源漏区的注入。这后两步对于多晶硅来说是对准的。对准工艺是得到小加工公差的关键。

材料的附加

有许多方法可以把材料附加到硅片上。可以把硅片放在炉子中进行氧化。对于大多数用炉子加热的工序和清洗工序，要把硅片放在“舟”中加工。一个舟可以装 10 至 25 片。在进行氧化时，Si 变为 SiO_2 。因为二者的体积比约为 1:2，所以 SiO_2 要超出原始 Si 表面，在 Si 表面以下部分和超过表面的量大体相等。 SiO_2 是无定形的。 SiO_2 可在 850°C 至 1000°C 间的任一温度下生长。在 900°C 下大约用 5.5min 就可以生长 400 Å SiO_2 ，而在同样温度下要用 190min 才能生长 4000 Å SiO_2 ，其生长速率随厚度的增加而下降，这是因为氧必须通过扩散穿过 SiO_2 才能到达 Si 表面。由于扩散方程具有时间平方的性质，所以要生长非常厚的 SiO_2 就要用很长的时间。在 1000°C 下生长 4000 Å SiO_2 ，大约要用 48min。因为氧化生长是高温工艺，所以它会改变 Si/ SiO_2 界面附近的杂质浓度。关键问题是杂质倾向于留在 Si 中还是留在 SiO_2 中。例如， SiO_2 往 Si 表面

内生长时，硼在硅表面耗尽，而在界面处的硅中，磷的浓度却是增加的。在 SiO_2 中的杂质对我们不大重要。可以用 Si_3N_4 来掩蔽硅表面的氧化。

可用在杂质气氛中扩散的方法向硅表面掺杂。精确的杂质分布取决于杂质源的类型，在体内，杂质分布按 $\exp(-x/\sqrt{Dt})$ 规律下降。其中 x 是硅片内离表面的距离， D 是扩散系数， t 是扩散时间。扩散系数是温度的强烈函数。硼和磷的扩散系数大体相等，而砷的扩散系数大约比磷小一个数量级。这就意味着如果其它条件相同，磷的扩散深度大约是砷的扩散深度的 π 倍¹⁾。可用 SiO_2 或 Si_3N_4 掩蔽这些杂质的扩散。最大掺杂浓度决定于固溶度，并影响到电阻的最小值和掺杂分布。从图 3.5 可以看出在一般情况下，横向扩散大约为纵向扩散的 70%。

离子注入是另一种掺杂方法。几乎可以用任何一种材料（包括光致抗蚀剂）掩蔽离子注入。这项低温工艺技术在 2.3 节中曾作过广泛的讨论。离子注入以后需要一步退火工艺，以便恢复晶格的损伤和使杂质电激活。退火温度范围是从 900°C 至 1000°C 。这么高的温度范围或许是必须的，以便给离子足够的能量，使它们运动到晶格中的适当位置，并且成为电激活的。有时候我们不希望某些杂质扩散，所以研究了一些快速退火技术。通过激光、电子束或辐射加热给硅增加能量，使得晶格温度只升高 10 至 20 秒钟。离子注入有重要的杂质二级效应，这对于硼特别严重。注入的硼沿着硅晶面穿行的能力很强，沟道效应会使最小结深加大。

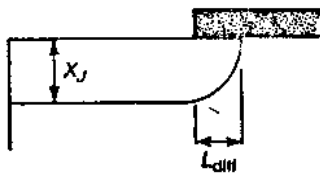


图 3.5 在角区的横向扩散。
一般情况下 $L_{diff} = 0.7x_j$

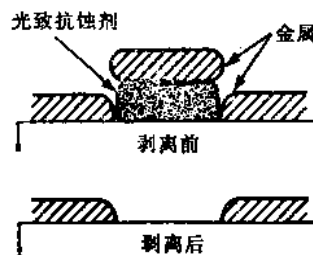


图 3.6 金属的剥离。当抗蚀剂被去掉时，沉积在抗蚀剂上面的金属就被漂洗掉了。采用这种方法，金属沉积对台阶的覆盖肯定很差

Liu 和 Oldham [Liu 83] 曾讨论过这个问题和某些改进技术。离子注入可以用光致抗蚀剂掩蔽，并且可以说是一种低温工艺。这两点使得离子注入成为 VLSI 制造的重要的工艺之一。这种工艺的主要限制是离子注入能够实现的注入深度小于数微米。由于这个原因，要继续采用长时间的扩散来制造深结。如 CMOS 工艺中的阱工艺。当然，要扩散的杂质可以用离子注入引入。

薄膜淀积是用来附加材料的又一种常用技术。蒸发法是加热被淀积的材料直至使它蒸发，然后它就像飘雪花一样飘落而凝聚在冷的硅片表面。用这种工艺所得的台阶覆盖是很差的。在剥离 (lift off) 技术中，蒸发可以方便地用作某些材料（例如铝）的选择性淀积。采用剥离方法时，对那些不需要材料的地方用光致抗蚀剂掩蔽，淀积后（当然所有地方都沉积上了材料）把光致抗蚀剂溶解掉，在它上面的那些淀积的材料也就被同时漂洗掉了，因而只有淀积在没有抗蚀剂的地方的材料才保留下来。图 3.6 为这种技术的示意图。

1) 对 π 取大值则 $\pi \approx \sqrt{10}$ 。

溅射是又一个通用的附加材料技术。事实上,大多数情况下铝是用溅射淀积的。用氩气束流轰击要附加的靶材料,材料落得到处都是,硅片上也落满了被淀积的靶材料。加工时硅片要加上偏压以便吸引靶材料的离子,并使二次电子不会打在硅片上而引起硅片发热。

电镀技术是把要附加的材料和硅片浸放到电解液中,强迫电流按某个方向流动,以便使离子吸到硅片上去。对于形成非常厚的金属层,例如,束射式引线器件中的金属层,电镀则是一种十分重要的技术。比如在束射式引线器件中,从芯片往外伸展的引线是作为芯片的一个部分加工出来的。在一般情况下,由于枝状生长使得电镀的金属非常粗糙。

前面阐述的外延生长技术是化学汽相淀积(CVD)的一种特例。由于参加反应的CVD气体是与芯片处在同一容器中,所以台阶覆盖很好。 Si , SiO_2 , Si_3N_4 和许多其它材料都可以用这种方法淀积。典型的CVD温度范围是从 380°C 至 600°C 。LPCVD(低压CVD)是CVD的一种特殊情况。有些LPCVD技术可以提供优良而均匀一致的沉积层(对于剥离技术而言却需要各向异性的沉积层,在良好的台阶覆盖成为问题时则需要均匀淀积层)。为了获得最低的CVD温度,往往要引用光能以促进化学反应。

加工好的硅片和管芯封装

图3.7示出了一块加工好的硅片。硅片加工完毕后,必然要通过封装工序。这包括用探针检测硅片、管芯分离、粘片、压焊和包封。

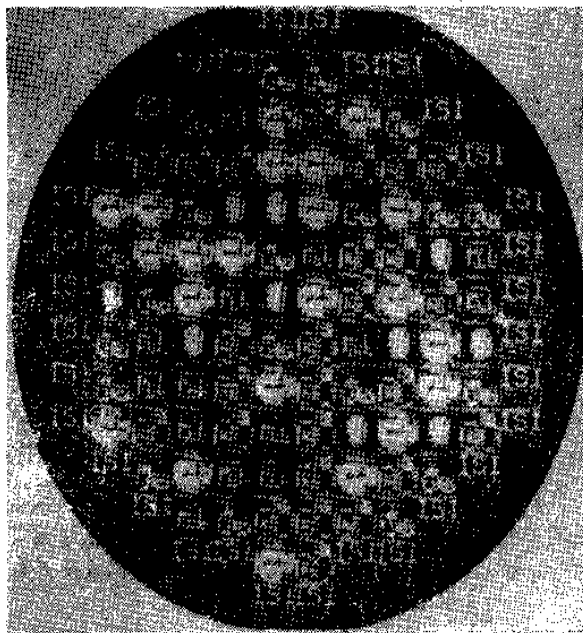


图3.7 完工后的含多个管芯的硅片。5 μm P阱CMOS工艺,采用了双层多晶硅和单层金属连线

硅片在自动测试系统上进行检测,测试图案通过探针加到管芯上。用来确定被测器件是否工作正常,要进行大量的测量,它所耗费的时间,会对大大提高最终产品的成本产生巨大影响。当发现了坏的芯片时,就滴上一滴墨水标志出来。

硅片进行了测试和用墨水标出之后，就用金刚石锯把它们切开，或者用金刚石针划开。在划片之后，把硅片放在一种薄膜上使其裂开。这个技术与切割玻璃类似。在图 3.8 所示的就是一些芯片。芯片可用高温焊料（常用含金的低熔点合金）¹⁾ 焊到管壳中去，更普通的办法是用环氧树脂封装。

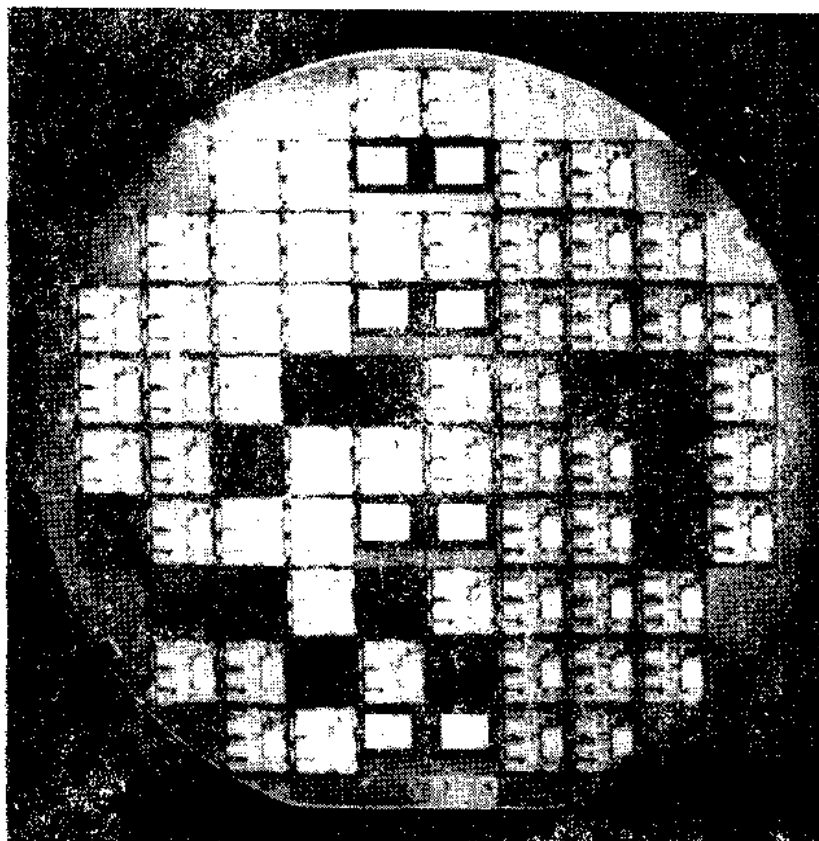


图 3.8 各种未压焊芯片

压焊是把芯片连到外部的最普通方法。常用的方法有两种。一是在 240°C 左右用金丝热压焊，在金与金压焊时这个方法是最成功的。它也可以在低温条件下用于铝压焊块的焊接。焊点的质量很高。压焊分两步，首先把焊丝压焊到芯片压焊块上，再把丝焊到管壳上。第二种压焊方法是超声焊，在这种情况下，采用直径为 20—30 μm 的铝丝。对于 MOS LSI 来说，超声焊是更为通用的方法。多数半导体工序是整片或整舟地进行的，压焊则例外，所以这一步是很昂贵的。图 3.9 中给出了一个封盖前压焊好的芯片。

IBM 广泛采用的另外一种焊接方法是凸点焊法。把焊接凸点电镀在压焊块上，再藉助于凸点把芯片焊接到管壳上。为了补偿热膨胀效应，往往将焊点安排在靠近芯片的中心部位。

集成电路的管壳可以是金属的、陶瓷的或者是塑料的。金属管壳具有最佳的热特性，塑料的热特性最差，但却是最便宜的。一些公司使用陶瓷芯片载体，这种载体中封装了多个互相连接的管芯。

1) 把芯片放到预先熔融的焊料上边，并用超声振动芯片以便冲破焊料的表皮层。

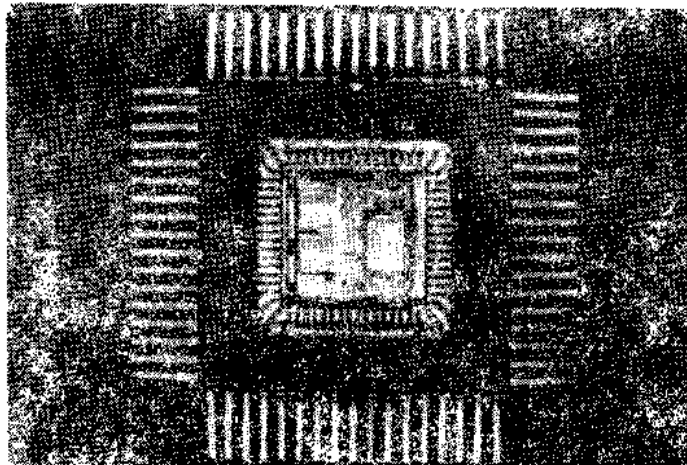


图 3.9 已压焊的芯片

3.2 两种基本的MOS 加工工艺

可以用各种方式把上节讨论的技术组合起来，用以制造 MOS 集成电路。我们将从硅单晶片开始介绍制造 MOS IC 的加工工艺。典型的硅片是直径大于或近似于 100mm 的圆片，其厚度为 300 μ m。硅片表面是镜面抛光的。因为 Si 的加工对沾污十分敏感，所以整个加工过程要在超净环境中进行。在许多加工工序之间，都要交替地进行清洗以去除片子上的沾污。下面将首先讨论基本的 MOS 加工工艺，然后再讨论 n 阱 CMOS 工艺。

基本的 nMOS 加工工艺

制造 n MOS IC 的第一步工序往往是把 P 型硅片放进氧化炉中，以便生成一薄层 (300 Å) 的高质量 SiO_2 缓冲层。在局部氧化工艺中，下一步是淀积一层薄的 Si_3N_4 。如果不用 SiO_2 缓冲层， Si_3N_4 就会带来表面态问题。这层 SiO_2 称作消除应力氧化层或者叫作缓冲层。这时片子已经准备好了可以进行第一步光刻。通过旋转把光致抗蚀剂涂覆在硅片上，随后用第一层掩模进行曝光，如图 3.10 所示。这一层是有源区掩模（也叫作 n^+ 区或扩散区掩模），它所限定的区域就是最后的源、漏和沟道区以及布线地道。对光致抗蚀剂显影并用光来掩蔽 Si_3N_4 不被等离子刻蚀。把硅片放进离子注入机内进行场区硼离子注入，如图 3.11 所示。光致抗蚀剂和 Si_3N_4 屏蔽离子不对沟道区、源区和漏区掺杂。随后去掉光致抗蚀剂。

硅片再放回到氧化炉中，并生长厚的场区氧化层。 Si_3N_4 层阻止周围的氧到达硅的表面。作为一级近似，只有在没有 Si_3N_4 的区域才生长厚的场区氧化层，如图 3.12 所示。然而存在一些非常重要的二级效应。强场注入的杂质会部分地进入到被掩蔽的区域。这是引起晶体管 ΔW 效应的一部分原因，在 2.3 节中曾经讨论过这种效应。另一个二级效应是 SiO_2 会从边上侵蚀到 Si_3N_4 区中。由于一个 SiO_2 分子所占据的体积近似为一个 Si 分子所占体积的两倍，因此会产生如图 3.12 中所示的鸟嘴结构。鸟嘴的缓慢斜坡，使其淀积材料能够有良好的台阶覆盖。但在另一方面，鸟嘴的大小对于 4 μ m 以下的工艺又是一个严重的问题。生长 SiO_2 的高温工序所引起的硼离子再分布是值得注意的。

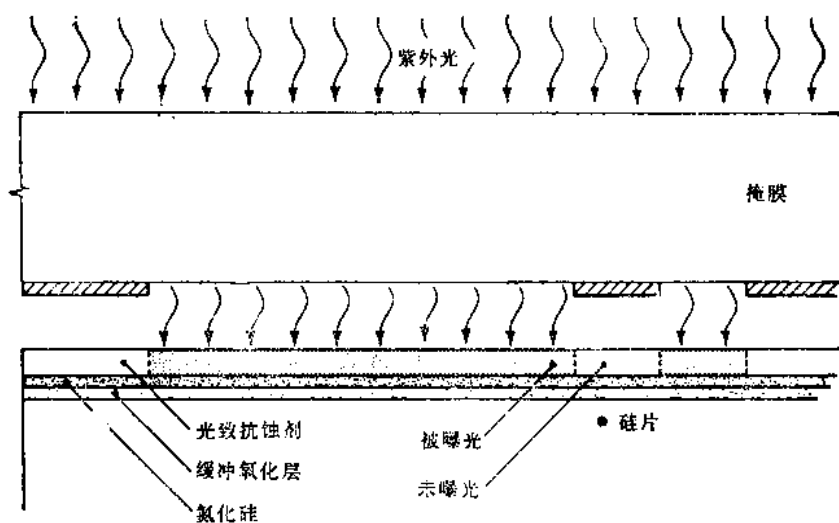


图3.10 用紫外光透过有源区掩模进行曝光

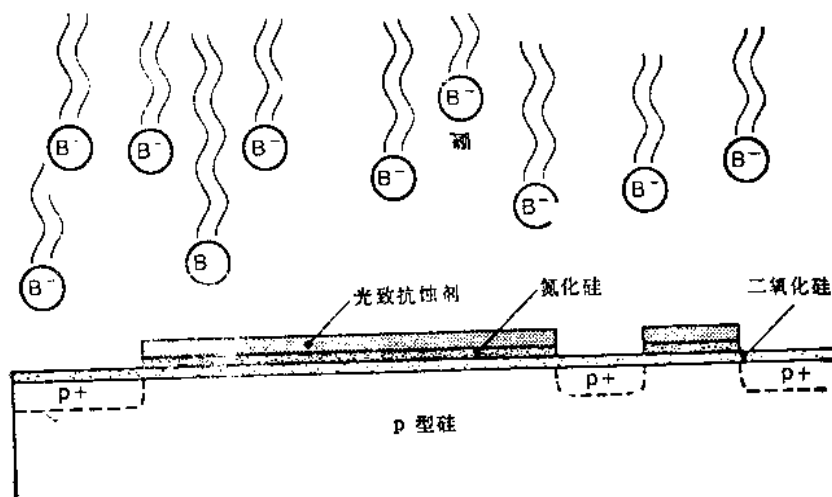


图3.11 用有源区掩模确定的光致抗蚀剂图形来掩蔽场区的硼注入

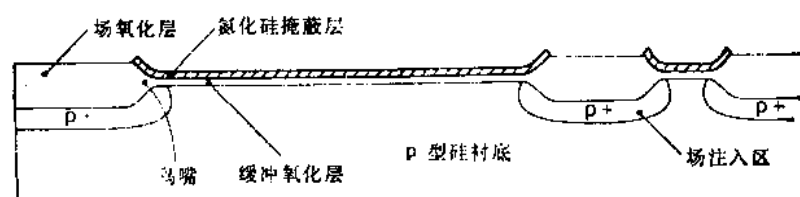


图3.12 用氮化硅掩蔽场氧化层的生长。氮化硅下面的缓冲氧化层有助于降低硅片表面的应力。在场氧化之后要去掉氮化硅

显然，如果打算得到非常精细的结构，则不希望前步工艺中引入的杂质产生移动。这就是倾向于低温加工的原因。

随后要去掉 Si_3N_4 、缓冲 SiO_2 层和一小部分场区氧化层，接着再生长一个非常薄的 SiO_2 层，这层就是栅氧化层。由于 SiO_2 要吃进到 Si 中去，所以栅下边的 Si/SiO_2 界面

应是没有受到污染的材料。事实上，在许多工艺中往往要生长和剥除一些栅氧化层以保证纯度。像场氧和栅氧化层这样的热生长氧化层，具有很高的纯度，它们具有较高的电场击穿阈值电压。热生长氧化层往往比化学汽相淀积的氧化层更抗腐蚀。为了工作可靠，栅氧化层的工作电场不要超过 2MV/cm 。

其后的几步工序是调整不同类型晶体管的阈值电压。把厚光致抗蚀剂涂布到硅片上并用离子注入掩模进行曝光。在显影之后，用厚的光致抗蚀剂层作为离子注入的掩模。注入的离子可以穿过薄的栅氧化层，该氧化层保护了硅表面不受沾污。在有源区和注入掩模重叠的区域才能注入离子。也可以用离子注入调整沟道区中的杂质分布。在先进的工艺中还用离子注入控制靠近晶体管处的源漏区掺杂分布，这样就可以减小短沟道效应和热电子效应。

把光致抗蚀剂又一次涂布到硅片上，用隐埋接触孔掩模定出多晶硅和 n^+ 区需要接触的部位，显影并用常规方法腐蚀出埋孔。腐蚀埋孔时只需要去掉一小部分栅区中的薄氧化层。只是在有源区和埋孔掩模重叠的区域才能腐蚀出氧化层上的接触孔。

应当再次指出，存在很多种前述典型工艺流程的可能变种。例如，对下一步工序，我们可以通过化学汽相淀积一层含重掺杂磷的多晶硅，也可以代之以淀积不掺磷的多晶硅，然后在腐蚀之后再与形成源漏区时同时掺杂。磷用于多晶硅掺杂是因为能得到较低的极小电阻率。用多晶硅掩模曝光并腐蚀出多晶硅图形，如图 3.13 所示。多晶硅可用作栅氧化层的掩蔽层，凡是没有多晶硅保护的地方，栅氧化层都被去掉，如图 3.14 所示，因而实现了栅与多晶硅的自对准。

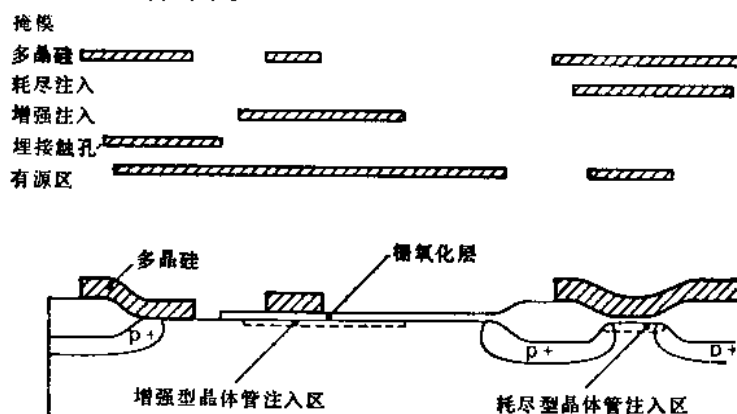


图 3.13 经过栅氧化层生长，阈值电压调整注入，埋接触孔腐蚀，多晶硅淀积，掺杂和腐蚀之后的 nMOS 硅片。这些工序的掩模图画在片子的上方

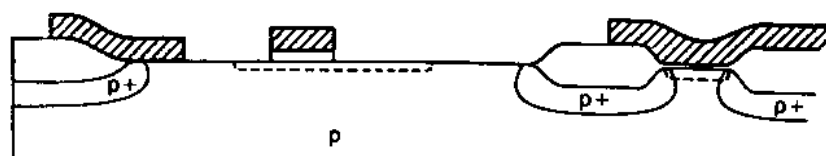


图 3.14 用多晶硅作为掩模腐蚀栅氧化层。这就实现了硅栅 nMOS 的“自对准”

在下一步工序中，用高剂量的砷注入得到源区和漏区。采用砷是因为砷的扩散系数比磷低，因此，对于精细尺寸的加工，由于砷的进一步推进很少，从而可以产生浅结和限制横向扩散。

随后将硅片进行微弱再氧化。这一步同时完成了好几件事：经退火去除离子注入引起的损伤；使杂质原子扩散，多晶硅中的磷扩散到埋孔下的硅中去，在相应区中得到较深的 $p-n$ 结；在多晶硅和源漏区上形成一层薄的完好的氧化层。源和漏区是高浓度 n 注入区，只有在有源区掩模和多晶硅掩模及版的交叠区 n 才能注进去。与 Si_3N_4 掩蔽工艺会产生鸟嘴的物理原因一样，靠近源和漏区的栅氧化层会变得略微厚些。因为这个区域是在晶体管的横向扩散 n^+ 区上面，所以这步工序有助于降低栅交叠电容（在先进的工艺中氧化时间较短，所以这个效应不十分显著）。到此，我们得到图 3.15 所示的晶体管。后面必须进行的工序就是连线。

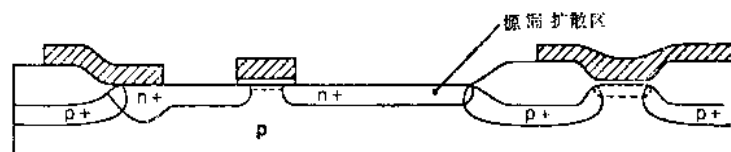


图 3.15 源漏区扩散时，多晶硅将微弱氧化，埋孔区杂质将略向衬底推进， n^+ 区往往是注入的，与此同时，磷也注入到多晶硅中

然后用化学汽相淀积更厚的氧化层。这一步往往用磷硅玻璃代替纯 SiO_2 ，以便使芯片钝化。含有 P_2O_5 的 SiO_2 有助于使钠离子远离 Si 表面，因此能阻止由于 Na 沾污引起的平带电压漂移，从而稳定了器件的阈值电压。这一点是通过吸取或键合钠离子而实现的。然后再将硅片加热使 CVD 的玻璃回流，以便为后续工艺提供平滑的表面。玻璃中的磷有助于回流工艺。对于浅结工艺，由于回流会加深结深，因而要避免采用，然而这样就会造成台阶覆盖不良。再下一步就是形成接触孔（有时回流工序是在接触孔腐蚀以后进行的）。在 n^+ 区和多晶硅上要腐蚀出接触孔。注意，这些接触孔并不需要穿透热生长的场区厚氧化层。

对于先进工艺，接触孔的接触电阻和占用面积都越来越成为重要问题了，所以倾向于采用自对准的接触工艺。图 3.16 和图 3.17 说明了自对准接触工艺的一些要点。在图 3.16 中，我们看到一种制作多晶硅自对准接触工艺 [Rideout 79, 80, Sakanoto 80,

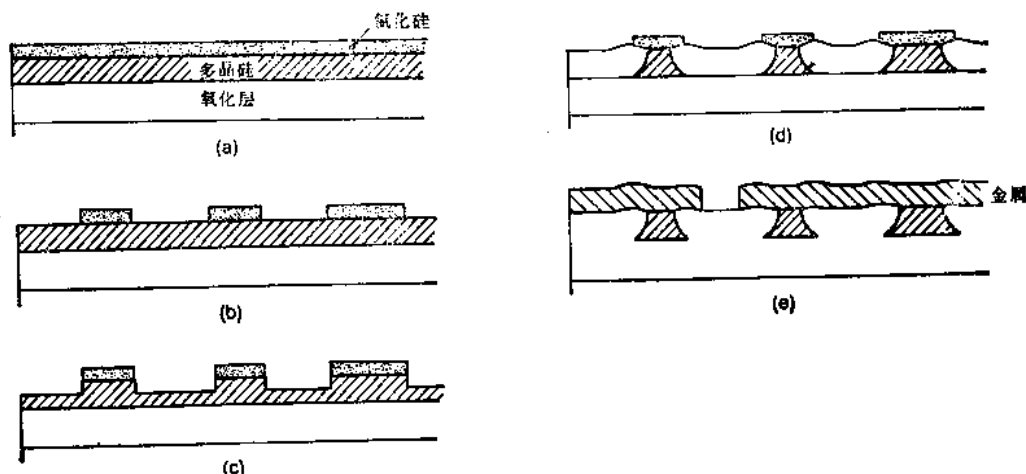


图 3.16 一种金属与多晶硅的自对准接触孔工艺。(a) Si_3N_4 在多晶硅上，多晶硅在 SiO_2 层上；(b) 形成 Si_3N_4 图形；(c) 腐蚀一半 Si_3N_4 ；(d) 多晶硅热氧化；(e) Si_3N_4 去除和在需要的地方制作金属接触

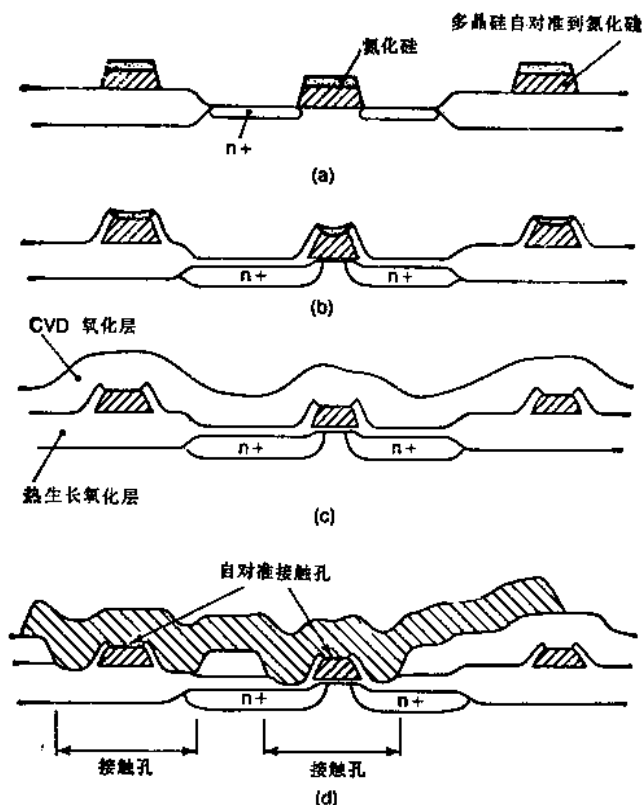


图 3.17 形成自对准接触的第二种技术。这种技术的关键在于热生长的氧化层必须比CVD氧化层的抗腐蚀能力强得多。(a) 用 Si_3N_4 掩蔽多晶硅的腐蚀，在一般情况下，用它定出扩散区和栅氧化区；(b) 硅片热氧化；(c) 淀积CVD氧化层；(d) 在CVD氧化层上腐蚀出超尺寸的孔。在达到热氧化层时，腐蚀速率就减慢了。再淀积金属和刻出图形

Muramoto 78)。这项技术的关键是对不需要接触的多晶硅进行选择性氧化。去除 Si_3N_4 层后，就露出想要制作接触的多晶硅。图 3.17 是一种有关用来接触多晶硅的技术，它可以在栅材料上形成接触。因为有些 CVD 氧化层的腐蚀可以比热生长氧化层的腐蚀快一个数量级，所以这个技术是成功的。于是可以在多晶硅上开一个大的接触窗口，而却能够保证不会腐蚀到衬底或扩散的硅区。这些技术的共同问题是它们都要求高温加工。

然后把铝溅射到硅片上，通过金属掩模刻出图形，这个铝层形成了第三层连线。最后加上作为保护层的玻璃覆盖层。通常用 SiO_2 上盖有 Si_3N_4 的复合层作为这一层玻璃覆盖层。图 3.18 是进行到这一步工序时的剖面图。在这个玻璃覆盖层上腐蚀出一些大的孔，就可以露出压焊块。基本 nMOS 工艺步骤可以概括如下：

- 生长缓冲 SiO_2 层并淀积 Si_3N_4 。旋转涂覆光致抗蚀剂。
- 用有源区掩模曝光，定出最终的沟道区和 n^+ 区。对光致抗蚀剂显影并腐蚀 Si_3N_4 。
- 进行场区注入。然后剥除光致抗蚀剂。
- 用 Si_3N_4 作为掩模生长场区 SiO_2 层。去除残留的 Si_3N_4 。
- 生长栅氧化层。用隐埋接触孔掩模刻出并腐蚀出埋孔。
- 用阈值调整掩模定出不同类型的晶体管。阈值调整是通过离子注入实现的，注入

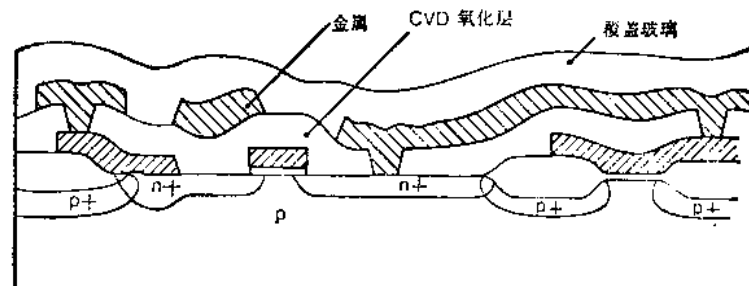


图3.18 具有普通接触孔的最终MOS硅片。虽然这是平面工艺，但是它的顶部并不十分平整

时用光致抗蚀剂作为掩模。注入离子可穿过栅氧化层但不能穿过厚得多的场区氧化层。

- 淀积多晶硅并用多晶硅掩模刻出图形。再用多晶硅作为腐蚀栅 SiO_2 层的掩模。
- 注入砷形成 n^+ 区。源/漏区对于多晶硅是自对准的。将片子再氧化以便注入的离子激活并推进埋孔区的 $p-n$ 结。
- 淀积 CVD 氧化层。用接触孔掩模刻出金属接触孔，腐蚀出接触孔。
- 淀积金属。用金属掩模进行光刻，腐蚀金属层。
- 淀积玻璃覆盖层。用玻璃覆盖层的掩模进行光刻。腐蚀出作为压焊用的窗口。
- 用探针检测硅片，用墨水标注坏的芯片。
- 切割和分离出芯片。
- 装架、压焊、封装和再测试。

例 3.1 内部含有埋接触的剖面图。

图 3.19 是在耗尽型上拉管中使用内部埋接触的俯视图和剖面图。因为该晶体管的栅长不好控制，如果它的电导要求严格则必须把它画的非常长才行。在降压电路结构中这种用法是普遍的。请注意在埋孔下边磷的深扩散区。

n 阱 CMOS 工艺

前述 n MOS 工艺和 n 阱 CMOS 工艺之间的基本区别在于第一步工序。在 CMOS 工艺中，第一步工序是确定并扩散出阱区，如图 3.20 所示。 SiO_2 或 Si_3N_4 层可用作扩散的掩模。因为多晶硅要么全是 n 型的、要么全是 p 型的，所以，只有一种型号的接触孔才是欧姆接触，这就是为什么在此往往不用隐埋接触孔的缘故¹⁾。在大于 $2\mu\text{m}$ 左右的 CMOS 工艺中，也很少使用场区注入。在阈值电压调整注入之后淀积多晶硅，然后掺杂并光刻出图形。用 n^+ 区掩模确定出有源区，该区将注入砷。图 3.21 是 n^+ 区和 p^+ 区扩散前后的 CMOS 芯片。可用光致抗蚀剂作为掩模。然后用 p^+ 区掩模刻出 p^+ 源区和漏区。可以用 n^+ 区掩模的反版作为 p^+ 区掩模，或者是用同一掩模但分别使用正性和负性光致抗蚀剂。在有些情况下，不用互为反版的掩模版，如果希望能制作肖特基势垒二极管的话²⁾ 就要求这样作。在 p^+ 区中， n^+ 多晶硅将被掺入反型杂质，因

1) 如果用 n^+ 多晶硅，在 n 阱中的隐埋接触将形成 $p-n$ 结（假定不使它们与阱区短路），实际上它们处于正向时才是有用的。用这种技术做成的反相器会在 p 沟器件和 n 沟器件之间存在一个正偏的二极管。从该 $p-n$ 结二极管注入的少数载流子会引起自锁效应。

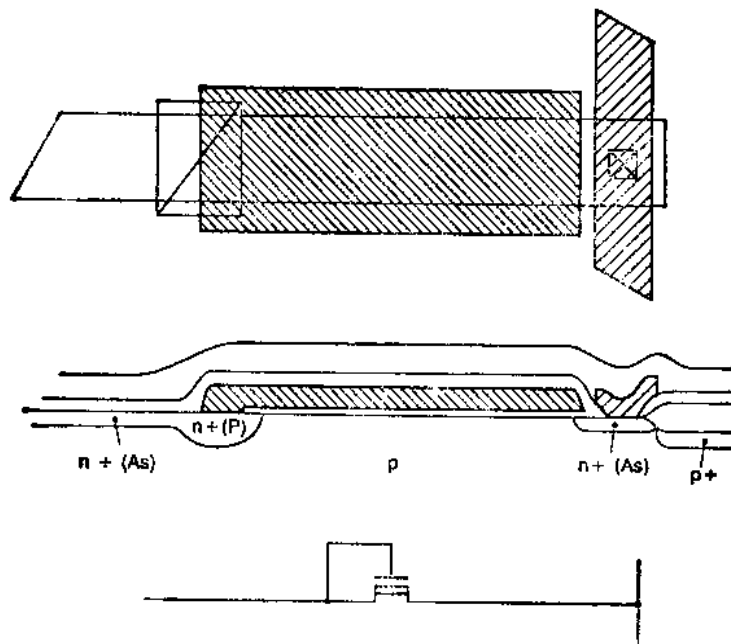


图 3.19 从顶视和侧视看到的内部埋接触孔。该晶体管栅长不好控制

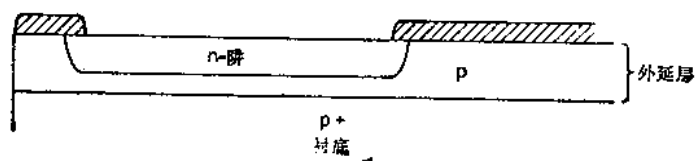


图 3.20 采用外延的CMOS工艺中的n阱扩散

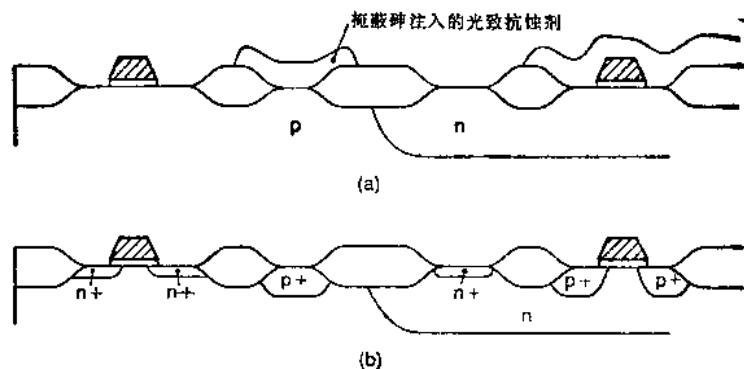


图 3.21 (a) 源/漏区注入之前；(b) 源/漏区注入后的剖面。所示光致抗蚀剂用作 n^+ 区注入的掩模。P 区多晶硅中掺有少量与 n 型相反的硼，使该多晶硅的电阻略有增加，但还保持为 n 型

为原来 n^+ 的剂量非常高所以多晶硅仍保持 n^+ 型，但是在这些区域内，多晶硅的电阻率就会略高一些。图 1.14 是一个典型 CMOS 版图的剖面图。其中 n^+ 源区和漏区比 p^+ 源区和漏区的结深要浅一些。这是由于二者掺入杂质的扩散系数不同，以及硼注入会产生注

- 1) 肖特基二极管是在半导体/金属界面处形成的一种二极管。Murphy 描述的工艺指出，在与形成欧姆结(线性电阻)相反的情况下，将得到肖特基势垒二极管。

入沟道效应所导致的结果。

有时采用多层金属布线。困难在于如果在各铝层之间用 SiO_2 作为介质,这时就不能用通常的方式进行回流,否则会使第一层铝熔化。有些制造者采用过特殊介质、特殊淀积技术和难熔金属来解决这个问题。有时采用聚酰亚胺作为介质,这是一种有机材料,适合于低温工艺,且具有很好的特性,能为后续工序提供非常平坦的表面。为 MOS 所用的难熔金属包括 Ta, W, Mo, TiSi_2 , TaSi_2 和 PtSi 。基本 CMOS 工艺的步骤可概括如下:

- 生长 SiO_2 层。旋转涂覆光致抗蚀剂并用 n 阱掩模曝光。
- 对光致抗蚀剂显影并腐蚀 SiO_2 。剥除光致抗蚀剂。然后在没有被 SiO_2 掩蔽的地方扩散 n 阱。
- 生长缓冲 SiO_2 层并淀积 Si_3N_4 。旋转涂覆光致抗蚀剂。
- 用有源区掩模曝光,定出最终的沟道区和扩散区。对光致抗蚀剂显影并腐蚀 Si_3N_4 。
- 用 Si_3N_4 作掩模生长场区 SiO_2 层。去除残留的 Si_3N_4 。
- 生长栅 SiO_2 层。
- 用 p^+ 和 n^+ 掩模调整晶体管阈值电压。
- 淀积多晶硅并用多晶硅掩模刻出图形。再用多晶硅作为腐蚀栅 SiO_2 层的掩模。
- 用 n^+ 掩模定出源/漏区以便进行砷注入。还要定出接到衬底的 p^+ 连接区。
- 淀积 CVD 氧化层。用接触孔掩模刻出金属接触孔,腐蚀出接触孔。
- 淀积金属。用金属掩模进行光刻,腐蚀金属层。
- 淀积覆盖玻璃。用玻璃覆盖层的掩模进行光刻。腐蚀出作压焊用的窗口。
- 用探针检测硅片,用墨水点掉坏的芯片。
- 切割并选出芯片。
- 装架、压焊、封装和再测试。

3.3 设计规则

设计规则是安排掩模上几何图形构图时所用的一些强制性规定。这些规则是通过综合考虑电学性能和可靠性限制与加工能力之间的关系而产生的。“多晶硅与多晶硅之间的间距必须大于或等于 $2\mu\text{m}$ ”就是一条典型的设计规则。一组设计规则的复杂程度不等,可能从 Mead 和 Conway [Mead 80, Plates 2, 3] 只用两页纸描述的“ λ ”那么简单的设计规则,到描述文件大大超过 50 页那样复杂的设计规则,这取决于你所解释的特殊情况数目,取决于工艺特性描述的细致程度和设计思想。图 3.22 说明了一些简单的设计规则。

每一个设计规则都是某些现象的相互作用结果。一个正确的设计规则起码要考虑到光刻工艺的特征、腐蚀工艺的能力、定位公差和电学特性的限制。定位公差与各层掩模之间预期对中偏差的方差有关。若希望设计规则适合于计算机程序检查,在制定设计规则时就必须考虑到这一点。

由于光刻在芯片加工工艺中的核心作用,所以在确定设计规则时光刻的限制起着重

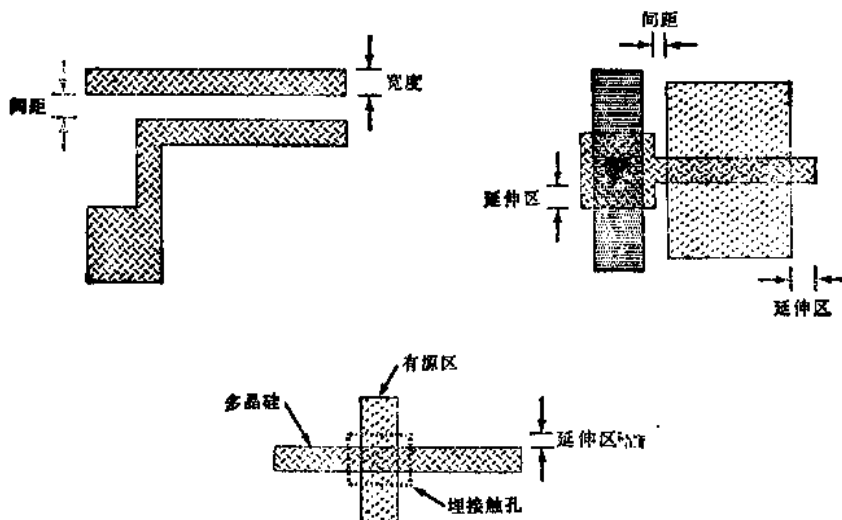


图3.22 一些典型的设计规则

要的作用。关键的衡量标准是光致抗蚀剂能分辨的最小几何尺寸，或者更本质地说，是小尺寸光致抗蚀剂的公差。这个特定的衡量标准从理论上说应同样地加到所有层次的规则中去。在1982年，用最先进的工艺在平整表面上加工的几何尺寸公差 3σ 小于 $0.5\mu\text{m}$ ，这里， σ 为标准偏差。一个复杂的问题是表面的平整程度可以对图像的质量起主要作用。由于半导体加工过程是一层一层地进行的，最上面一层可能会形成与其下层有关的不均匀形貌。在这种情况下，因为光线衍射所限制的成像景深大约为 $5\mu\text{m}$ ，上边一层的线条间距和线宽的设计规则就需要放松一些。这是促使后来采用多层抗蚀剂系统的原因之一。

与拓扑结构相关联的另一个效应是曝光用的紫外线反射。金属层要跨过许多凸起部位和凹陷部位。显然，在芯片内只有金属层而没有其它层的那些区域中，光刻质量会更好一些。在有許多台阶的部位，可能要采用一种其它规则（叫做反射规则）。回顾前述加工工序可知，在顶部高低不平的芯片表面上淀积金属。然后，在它上边覆盖光致抗蚀剂，这就很可能使穿过掩模透明部分照射到片子上的光线被凹陷处的边墙斜向反射。这样一来，就会使芯片的掩蔽区域曝光，如图3.23所示。反射规则就是试图通过改变金属间距设计规则来对付这种现象的。该规则中把金属间距规定为台阶种类和数目的函数。使用这种规则时，两条金属线的间距可能是 $3\mu\text{m}$ 或 $5\mu\text{m}$ ，而不再是 $4\mu\text{m}$ 。

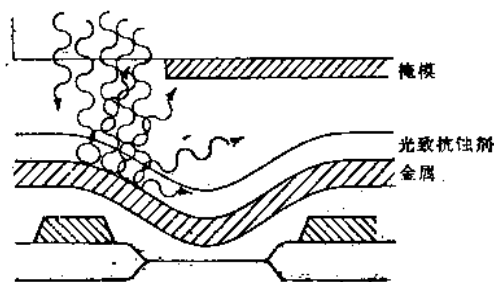


图3.23 金属边墙反射可引起被掩蔽的光致抗蚀剂意外曝光

开发一个CAD软件来检查这类设计规则是极端困难的。可以用一个盖在铝层上的抗反射层来避免反射问题，图3.24(a)为没有用抗反射覆盖层光刻出来的覆盖在水平多晶硅线上的金属线条，图3.24(b)是在采用抗反射覆盖层时刻出的这些线条。也可以对任一或全部掩模层特殊制定设计规则，这种特殊制定的设计规则要规定在某一掩

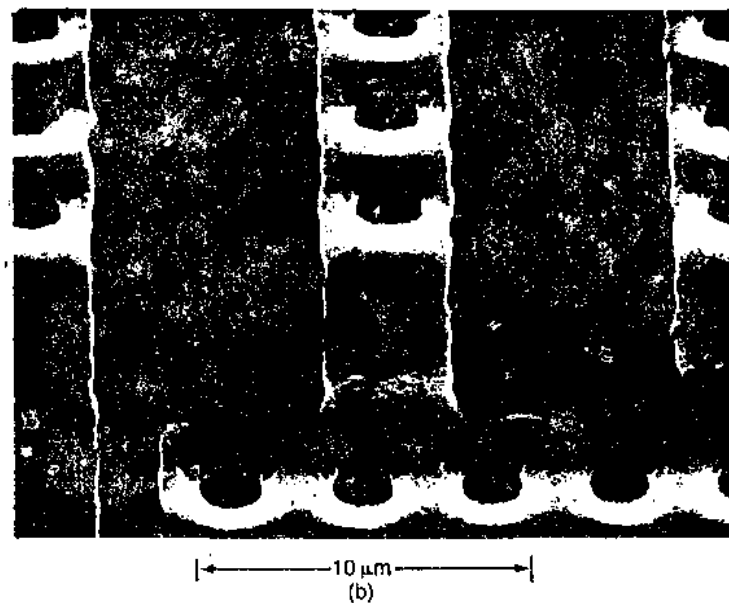
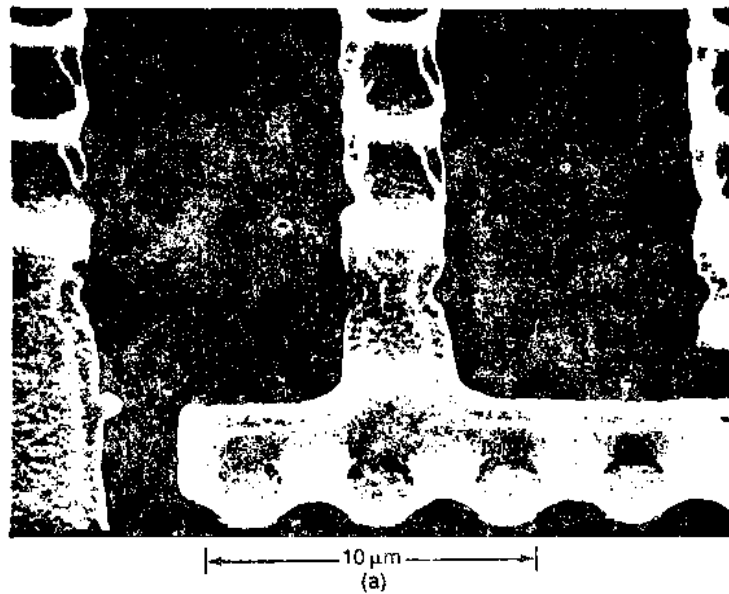


图3.24 金属线的电子显微照相。(a) 在没有抗反射层时刻出的金属线；(b) 在有抗反射层时刻出的金属线

模层上的几何尺寸必须大于某个最小值。这就便于查出掩模上明显的漏洞或错误（其次还可能会有小的几何图形在要害部位破损以及加大等问题）。

一旦在光致抗蚀剂中曝光图形并显影以后，就要对它下边的一层或多层材料进行腐蚀。这一步会对线宽尺寸引入附加的不确定性，最基本的要求是防止相邻连线短路和连线本身断路。传统的湿法化学腐蚀工艺是有选择性的，但却是各向同性的。因为它们从纵向腐蚀材料，也同样在横向进行腐蚀，所以难以严格控制尺寸，特别是对于厚膜的腐蚀更是如此。等离子和反应离子干法腐蚀工艺优先在垂直方向腐蚀，从而改进了对尺寸

的控制能力。在所有腐蚀方法中，对于较薄的膜，尺寸控制比较容易，同时也因材料的不同而稍有变化。

所有腐蚀工艺的共同困难是整个芯片上的腐蚀速率并不是均匀的，而是或多或少依赖于几何图形所在位置的情况。例如，附近什么都没有的一条细线，要比密集连线区中的细线腐蚀的快些。这是因为在腐蚀过程中，腐蚀剂的化学成分变化的结果。由于接触孔的尺寸对于版图密度是很关键的，而且是难以控制的，许多工艺师都把他们的工艺调整到刻蚀最小尺寸接触孔为最佳的状态。这样一来，比最小孔大得多的孔就呈现出模糊的边界而且不能控制。因此，要求接触孔最少有一个方向的尺寸是最小尺寸，有可能的话，在两个方向上都是最小尺寸。为了保证所有接触孔都能充分腐蚀，某些接触孔就会出现过腐蚀。

除第一层掩蔽膜以外，所有后续掩蔽膜都必须与以前的掩蔽层或工艺步骤对准。在大约 1980 年以前，这些对准都是靠操作人员用手工通过将一些称为对准符号的同心框对在一起实现的。最近，效果更好和重复性更佳的自动对准系统才付诸使用。在硅片步进技术中，每一步进图案都可达到芯片与芯片的对准，这个优点在对准技术中起了关键作用。

在拟定设计规则时，对准顺序的选定有着重大的影响。例如，现有掩蔽和工艺步骤 A、B 和 C。若 B 对 A 对准的公差为 $\pm 1.0\mu\text{m}$ ，C 也与 A 对准，且具有同样公差，则在最坏情况下 B 对 C 定位的公差为 $\pm 2\mu\text{m}$ 。如果我们选择 C 与 B 对准，则 C 与 A 对准就有相似的公差 ($\pm 2\mu\text{m}$)。实际上，公差是统计性的而不是代数相加。通常采用平方根的关系，即

$$R = R_0 \sqrt{n} \quad (3.1)$$

其中 R 是定位公差， R_0 是各层之间单个层次的定位公差，而 n 是间接对准层数。图 3.25 说明了典型 nMOS 和 CMOS 工艺的对准树。节点之间的距离表示 n ，这些数据也列在表 3.1 中。主要需判断是接触孔向多晶硅对准还是向有源区对准。

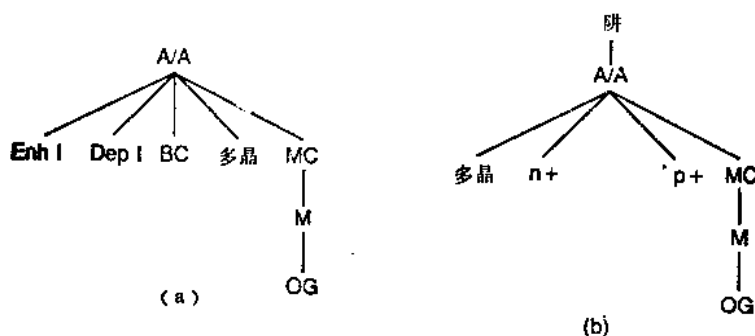


图 3.25 对准树。(a) 对于 nmos；(b) 对于 cmos。其中缩写 A/A 为有源区；Enh I 为增强注入；Dep I 为耗尽注入；BC 为埋接触；Poly 为多晶硅；MC 为金属接触孔；M 为金属；OG 为覆盖玻璃钝化层

材料特性及其相互作用带来了某些附加的限制。下边我们将讨论其中的一些问题。

各个层次都有一些相关的电学约束，使最小尺寸受到限制。金属层的约束最简单，其唯一的要求是相邻线条之间不能短路以及最细线条必须宽到足以保持其连通。

表3.1 MOS 对准表

n	nMOS对准表							
	A/A	ENH	Dep	BC	Poly	MC	M	OG
有源区	0	1	1	1	1	1	2	3
增强注入		0	2	2	2	2	3	4
耗尽注入			0	2	2	2	3	4
埋接触				0	2	2	3	4
多晶硅					0	2	3	4
金属接触						0	1	2
金属							0	1
覆盖玻璃								0

CMOS 对准表

n	W	A/A	poly	n ⁺	p ⁺	MC	M	OG
阱	0	1	2	2	2	2	3	4
有源区		0	1	1	1	1	2	3
多晶硅			0	2		2	3	4
n ⁺ 区				0	2	2	3	4
p ⁺ 区					0	2	3	4
金属接触						0	1	2
金属							0	1
覆盖玻璃								0

扩散区线条上所施加的约束是比较复杂的。选择性场区氧化所确定的边线会产生显著收缩。因此如果没有考虑 ΔW ，图上表示的最小线条就可能完全消失。扩散区线条的最小间距更为复杂，这是因为即使线条本身没有碰上，但是由于结的贯通会产生短路。

设计规则可以只涉及一层掩模，也可能与数层掩模有关。因为有源区掩模是第一层掩模，设计规则只对线条宽度和线条间距加以约束。离子注入掩模设计规则所考虑的是注入区与晶体管的间距，或者是超出晶体管的尺寸，即使是在掩模对准有偏差以及光致抗蚀剂曝光有误差时，也要确保应该注入的晶体管得到注入，而不应注入的晶体管不会被注入。

埋孔接触设计规则与离子注入设计规则是类似的，它要求刻出的埋孔处有埋接触孔，而其它地方则不能有这种接触。有些设计规则是复杂的，在设计时，多晶硅、有源区和埋孔接触在版图上没有交叠，然而由于错位和图形扩大，在最后的芯片上三者却有可能互相交叠起来。再者，由于磷的横向扩散大于砷，所以两个埋接触孔的间距要大于两个通常扩散区的距离。

多晶硅设计规则的宽度和间距往往是最小的。除此之外，设计规则中要有多晶硅与无关有源区的间距和栅超出有源区的尺寸。对于埋接触，规则中要有多晶硅与有源区的最小重叠面积。为了保证最小接触面积，设计规则中的有源区或多晶硅，或者二者同时在各个方向上超出重叠区。在多晶硅作为连线，而不是用它组成晶体管时，往往需要有多晶硅至有源区的最小间距规则。有了这条规则，就降低了多晶硅连线电容。如果多晶硅与相应有源区没有对准，以致于在部分多晶硅下存在薄氧化层，芯片上栅区面积就会大

大增加，所以这条规则也影响到成品率。

接触孔将金属层通过 CVD 氧化层连接到多晶硅或有源区，使这类接触孔处在正确的位置是很重要的，所以设计规则要包括有源区和多晶硅超出接触孔的尺寸。多晶硅必须与金属有足够的距离，否则二者就会短路。

金属接触孔不能靠近硅栅或在硅栅上面，这是因为金属铝会沿着晶粒间界渗透到多晶硅中去，并且从多晶硅穿通氧化层。铝是氧的吸收剂，这对于形成铝硅间的接触是有好处的，因为铝会穿通接触孔内必然存在的薄氧化层。另一方面，这个特性会引起栅短路。第二个问题是铝会改变栅材料的功函数。这些限制可以用阻挡层金属来解决。毗连接触 (butting contacts, 或称对接孔) 是一个有争议的方案，这种接触是用金属覆盖在稍有交叠的多晶硅和有源区接触孔上实现的，有人认为由于在靠近多晶/ n^+ 交接处，金属可能与该处衬底短路，而成为可靠性的隐患，在图 3.26 中说明了这一点。因为金属会溶解一些硅，而大量的金属会溶解更多的硅，所以与大块金属的接触也会出现问 题。解决的方法之一是将接触移到从大金属块延伸出的细条形区域内，另一个方法是向接触区提供一小块供铝“吃掉”的多晶硅。这些技术的说明在图 3.27 中。

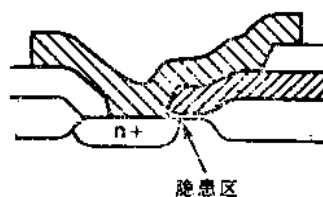


图 3.26 毗连接触。铝钉穿到靠近 n^+ /多晶交接处的衬底中，是长期可靠性的隐患

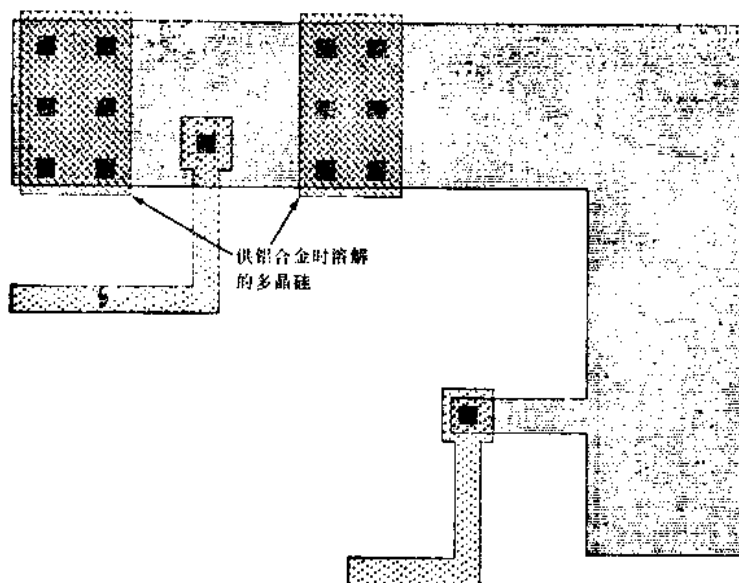


图 3.27 把有源区连接到大块金属的方法。硅溶解到铝中去会引起可靠性问题

金属的上面是保护层或玻璃覆盖层。压焊块窗口的尺寸和压焊块间距与压焊设备有关，往往还要规定其它掩模层必须离开压焊块的距离以免引起可能的损伤。

在 CMOS 电路中，有两条设计规则是为自锁效应而引入的。其一是 p 型衬底中 n^+ 区和 n 阱中 p^+ 区之间的间距；其二是应该离多远设置一个接到阱区的接点，每个阱区接点都通过一个 n^+ 区连到金属层并和 V_{DD} 相接。至于应该离多远接到阱区上，设计规则可以采取两种方式加以规定，第一种方式规定每经多少个微米就要安排一个阱区接

点；第二种方式中，规则规定在细长的阱区中要防止产生大的压降，即规定接点之间不应多于多少“方”的阱区。

除了电路以外，芯片中往往还要有划片记号和对准记号，并且也可能有一些测试结构、识别字句、名称字标和版权记号。

一个新工艺的设计规则是通过计算和经验结合得到的。在工艺不成熟时，随着工艺师优化工艺合格率，设计规则会不断地变更。采用这种工艺进行了几个设计之后，设计规则就必须稳定下来。

例3.2 设计规则计算。

作为导出设计规则的一个示例，我们计算与金属至多晶硅接触相关的设计规则。该工艺具有如下特点：

层 次	线宽公差	对准对象
多晶硅	$\pm 0.25\mu\text{m}/\text{边}$	有源区
金属	$\pm 0.40\mu\text{m}/\text{边}$	接触孔
接触孔	$+0.6$ -0.0 $\mu\text{m}/\text{边}$	有源区

上述接触孔线宽公差是对 $2\mu\text{m} \times 2\mu\text{m}$ 的接触孔而言的。请注意接触孔公差是不对称的。其对准公差 $R_0 = \pm 0.75\mu\text{m}$ 。

为了得到一个符合设计规则的接触孔，假定接触孔的边不能超过多晶硅。如果照这样做，我们必然面临台阶覆盖问题，如图 3.28(a) 所示，或者在更坏的情况下，会造成一个腐蚀坑，如图 3.28(b) 所示。工艺加工中的副产品会聚集到腐蚀坑（也叫作废料洞）中，所以会造成长期的可靠性隐患。为了使接触电阻不超过已定的最大值，我们假定多晶硅与金属的最小接触面积为 $3\mu\text{m}^2$ 。

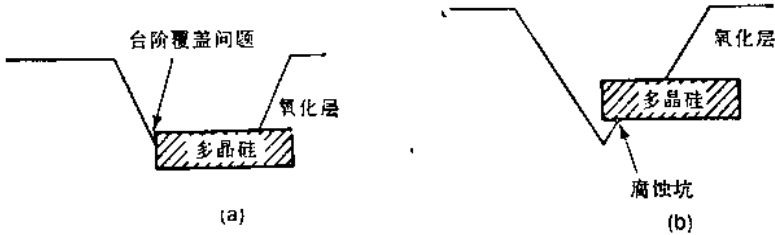


图 3.28 多晶硅与接触孔的对准失配和过腐蚀。(a) 台阶覆盖问题；(b) 腐蚀坑

我们首先考虑孔到多晶硅的定位值 R_{pc} 。因为二者都是与有源区对准， n 等于 2。所以

$$R_{pc} = \sqrt{2} R_0 = \pm 1.06 [\mu\text{m}]$$

最坏情况是多晶硅缩小而接触孔扩大。我们求多晶硅伸出接触孔的最小尺寸。加上最坏情况下的扩大、收缩和平移，有

$$1.06 + 0.25 + 0.6 = 1.91$$

这最小尺寸为 $1.91\mu\text{m}$ ，我们把它四舍五入为 $2\mu\text{m}$ 。图 3.29 说明了在最坏情况下孔对多晶硅的几何尺寸。

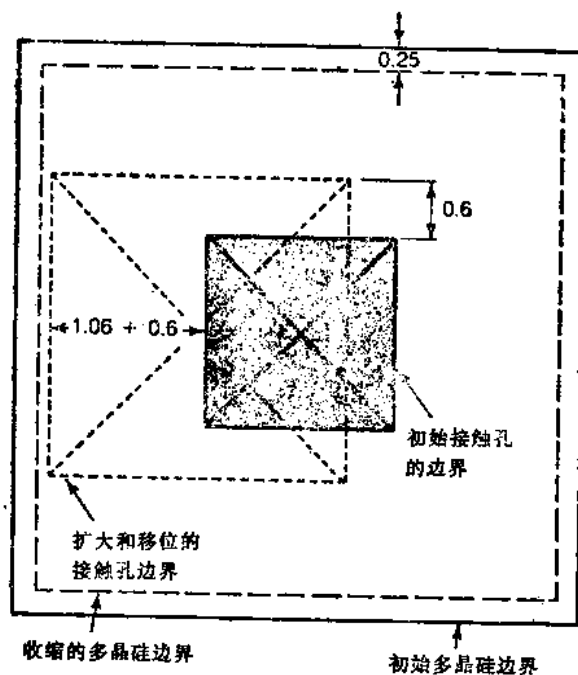


图3.29 实践表明理想几何尺寸。虚线是扩大、收缩和失配的组合的最坏情况

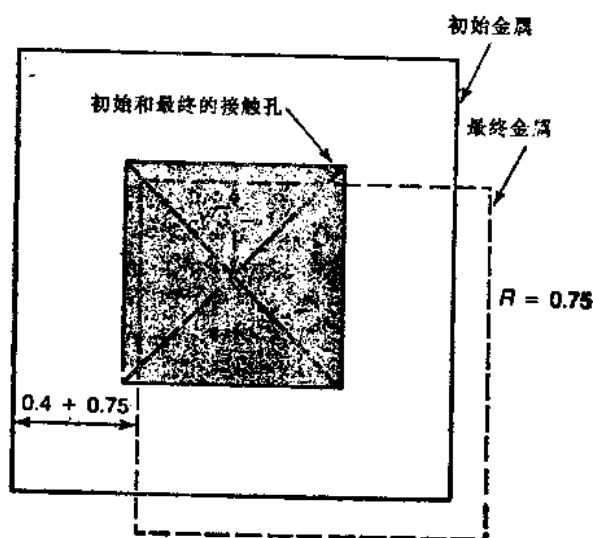


图3.30 金属覆盖孔的最坏情况。使接触面积最小

下边我们研究金属的设计规则。因为金属是与引线孔对准的，所以在这种情况下 $n = 1$ 。假定最坏的平移在 y 和 z 方向同时发生，则覆盖区的每一边长必定为 $\sqrt{3} \mu\text{m}$ 。这相当于从中心到边为 $0.73 \mu\text{m}$ ，如图 3.30 所示。把这些误差加起来，可有

$$0.73 + 0.75 + 0.40 = 1.88$$

从孔的中心到金属的边，应为 $1.88 \mu\text{m}$ ，我们把它四舍五入为 $2 \mu\text{m}$ 。因而金属覆盖在引线孔外的尺寸必须为 $1 \mu\text{m}$ 。如果要保证接触面积为初始值 $4 \mu\text{m}^2$ ，则这个尺寸应为

$$0.75 + 0.40 = 1.15 [\mu\text{m}]$$

现在, 假设 R_s 从 $\pm 0.75\mu\text{m}$ 下降到 $\pm 0.5\mu\text{m}$, 这不会改变多晶硅的覆盖规则问题, 但是得重新检查金属的覆盖规则。这时, $1.88\mu\text{m}$ 变为 $1.63\mu\text{m}$, 还是可以把它四舍五入为 $2\mu\text{m}$ 。再考查一个普遍情况, 即金属作为连线通过接触孔时对两条对边的覆盖都应大于 $1\mu\text{m}$ 。在这种情况下, 最坏情况的对准失配在垂直于连线的方向上。为了能有 $3\mu\text{m}^2$ 的接触面积, 就需要金属覆盖接触孔的面积为 $1.5\mu\text{m} \times 2\mu\text{m}$, 如图 3.31 所示。这样一来, 垂直于连线方向上的金属覆盖在孔外的最小尺寸就得有 $0.4\mu\text{m}$, 或舍入为 $0.5\mu\text{m}$ 。在允许连线金属为 $3\mu\text{m}$ 条宽的工艺技术中, 这一点将大大有利于版图设计。

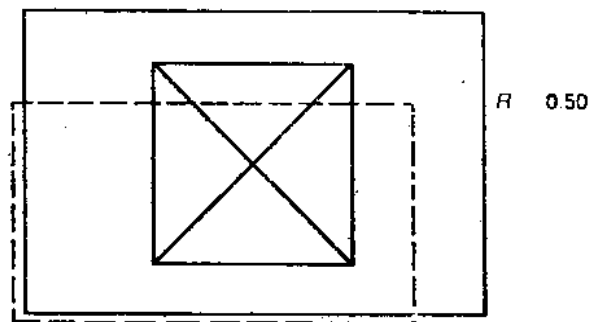


图 3.31 金属连线穿过接触孔的最坏情况

上面第二个例子说明了两点, 一, 设计规则制订得越细, 就可能使版图设计得越紧凑。如果所有设计规则都是 $1\mu\text{m}$ 尺寸的倍数, 则 $0.4\mu\text{m}$ 一定舍入为 $1\mu\text{m}$, 采用 $0.4\mu\text{m}$ 的设计规则, 就会毫无用处; 二, 设计规则的复杂性有些是看法不同的原因。例如, 在上述后一个例子中, 我们可能有两种金属覆盖的设计规则。首先是通过孔时, 在所有方向上金属都必须超出孔 $1\mu\text{m}$; 其次, 是通过孔时, 在两个相对方向上, 金属可以只超出孔 $0.5\mu\text{m}$, 只要在另两个方向上金属超出孔 $1.0\mu\text{m}$ 就可以。在设计规则中, 我们要不要包含这第二项规则? 如果在这种特殊情况下, 要求在长方向上超出孔的尺寸为 $1.5\mu\text{m}$, 又该怎么办呢?

例 3.3 设计规则的折衷。

在编制设计规则时, 要有许多折衷。有些参量比较重要, 所以必须坚持, 而另外一些则不然。以 CMOS 技术为例, 要考查阱内有源区与阱外有源区之间的间距, 一方面, 这个距离决定着自锁效应的敏感程度, 另一方面, 它又决定着版图的密度。在我们示例的工艺中, 这个最小距离设计得正好足够大, 使得在阱边缘处的各有源区之间能放下金属与多晶硅的接触孔就可以了。由于经常用到这种接触孔, 这对有源区进一步互相靠近是不利的。

3.4 成品率

成品率是表征特定制造工艺成功程度的最重要度量, 受设计者控制的芯片面积是影响成品率的主要参数 (假定遵守设计规则), 制造芯片的成本比芯片面积加大的速度还快。

有些特性和优值可与集成电路的制造联系起来。在这一节中, 我们只考虑四个因素。

一个集成电路的面积 A_0 ，给定芯片能工作的几率 p_0 ，每个硅片上合格器件的成品率 y_0 以及成本 C （归因于成品率损失）。一般情况下有

$$y \propto \frac{p}{A} \quad (3.2)$$

和

$$C \propto y^{-1} \quad (3.3)$$

为了得到一个简单的集成电路成品率模型，假定合格芯片的几率只与其上不存在点缺陷相关。我们进一步假定点缺陷的分布可用泊松统计描述。在这些假设下合格芯片的几率 p 具有如下形式：

$$p = e^{-A/A_{00}} \quad (3.4)$$

其中 A_{00} 为一常数，它取决于工艺质量并反比于缺陷密度。 A_{00} 是一个随机变量，它使分析变得相当复杂；它随线宽的增加按平方规律变化，且随苛刻工序的数目及缺陷密度而下降。在 A_{00} 变化大的时候， p 如下式所示[Price 70]：

$$p \approx \frac{1}{1 + A/A_{00}} \quad (3.5)$$

式 (3.5) 所预示的成品率远大于式 (3.4) 所示的结果，但看起来它与经验更为符合。

让我们考查如下情况，即面积为 A_0 的芯片不存在缺陷的几率为 p_0 ，相应的成品率为 y_0 ，成本为 C_0 。对于成本，我们有

$$C \propto \frac{A}{p} \propto A \left(1 + \frac{A}{A_{00}} \right) \quad (3.6)$$

或

$$\frac{C}{C_0} = \frac{A}{A_0} \left[p_0 + \frac{A}{A_0} (1 - p_0) \right] \quad (3.7)$$

在图 3.32 中，画出了归一化成本对归一化面积的关系，曲线对应于几个 p_0 值。从式 (3.7) 我们注意到，即便所有芯片都是无缺陷的，成本还是要随器件面积而上升，在

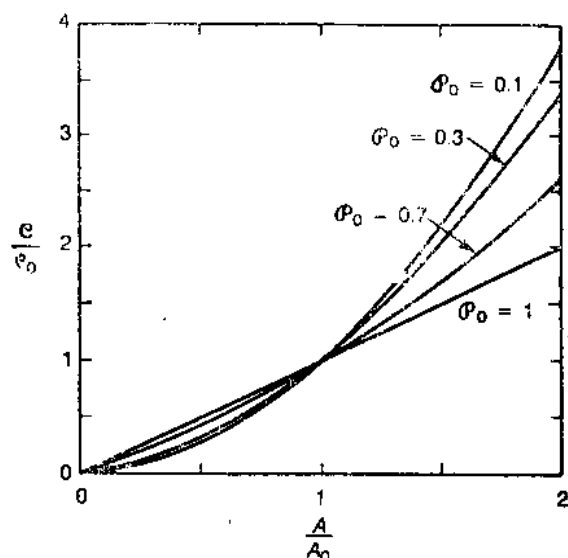


图 3.32 按成品率的价格模型，得到的归一化成本与面积的关系图

p_0 较小时, 成本随面积的上升更快。成本对面积的灵敏度为

$$\left. \frac{d(C/C_0)}{d(A/A_0)} \right|_{A=A_0} = 2 - p_0 \quad (3.8)$$

由此可见, 在 p_0 较小时成本上升的速度是较大 p_0 时的两倍(注意, 如果式(3.4)是有效的, 则式(3.8)将成为 $1 - \ln p_0$, 这是更为不利的函数形式)。

在实际问题中, 由于一些其它效应, 使这个芯片成品率的简单模型复杂化。这些效应包括有: 缺陷(如划伤)在空间上的相关性; 在 I/O 电路中失效率不是按比例地增高; 封装失效; 参量失配失效等等。在某种意义上, 缺陷之间有确定的关联将有利于成品率。即对于同样数目的缺陷, 如果在一个芯片上含有全部缺陷, 则在其它芯片上就不会再有缺陷, 这个理论对于成品率非常低的场合是不成立的。

缺陷处如果没有电路, 它就不会引起器件失效。这一点大大影响设计方法的选择。由于这个原因, 用单元法[Kang 81]和门阵列法设计电路时成品率往往比同样尺寸的用户定制芯片更高。采用容错改进了成品率, 但带来了另外的复杂性[Smith 81]。在恰当地采用容错技术时, 成品率和面积可能会同时增加。

3.5 版图设计技巧

版图设计是一种易于学习但难以掌握的技巧。设计规则是支配工艺的法则。图3.33给出了一个 nMOS 的版图设计, 它是按照附录C中的设计规则画出的。所用的各个设计规则是用规则号码标志的。在这个版图设计中, 省略了耗尽型和增强型的阈值调整掩模。尽管如此, 也必须遵守这些层次的设计规则。

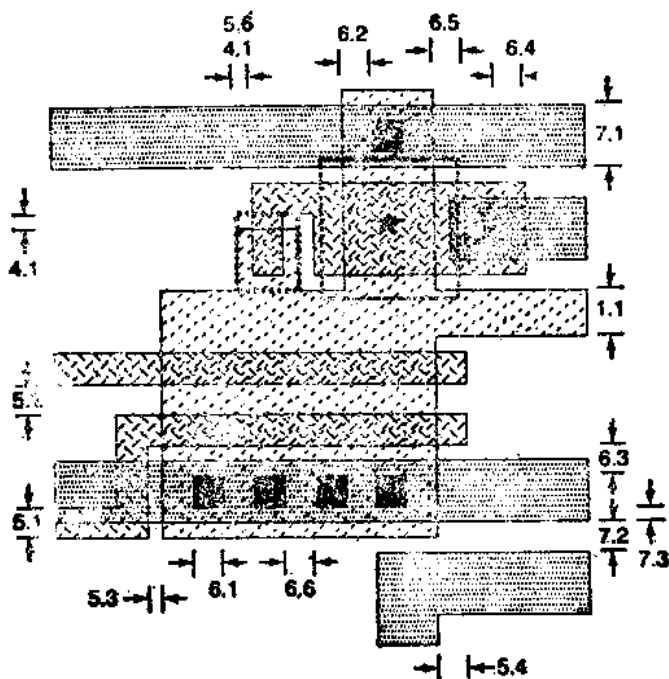


图3.33 说明附录C中nMOS设计规则的版图

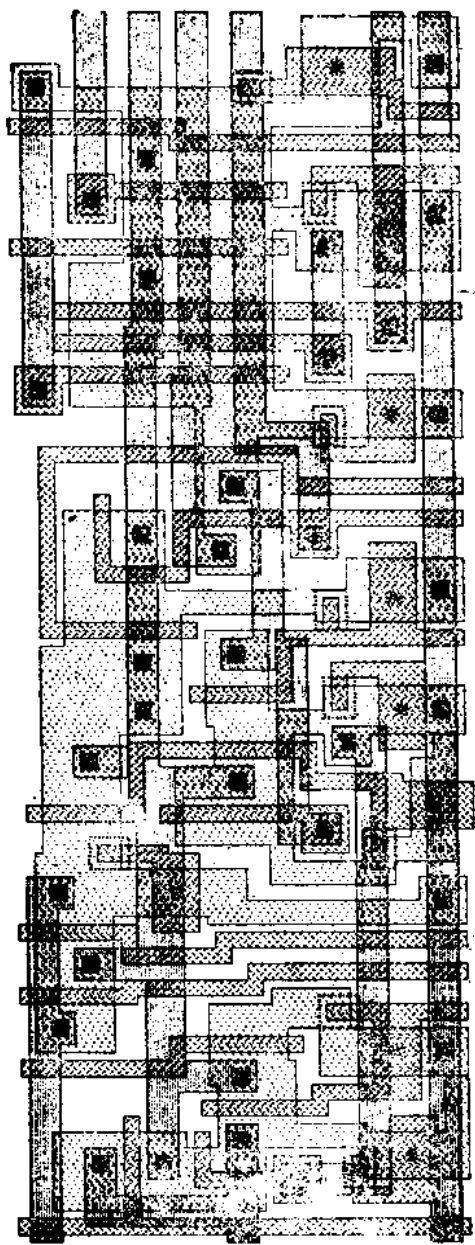


图3.35 nMOS随机逻辑的一小部分版图，
在随机逻辑设计领域，初学者和版图设计
专家之间的区别是十分明显的

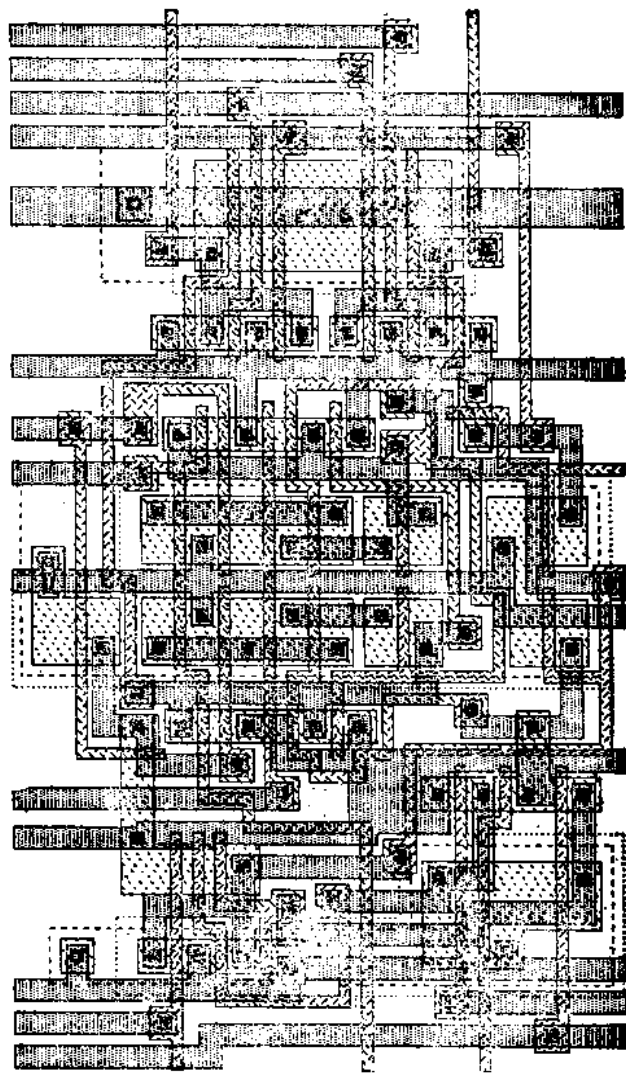


图3.36 CMOS的版图设计，它与图
3.35中的版图具有同样的逻辑功能

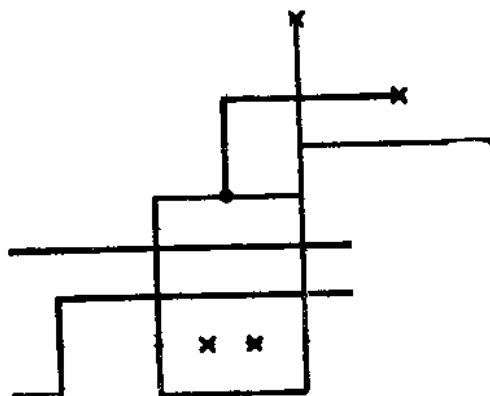


图3.34 说明图3.33中版图中的梗图 (stick diagram)。应该用彩色绘制梗图

对版图设计进行规划是很重要的，因为这种规划能很快地检验不同的设计。往往采用简化的符号来处理版图设计。这种简化符号称为梗符号，它是一种用来完整地说明版图的想像中的布局的表示图。梗设计图没有固定的规则。关键是把梗图画成能真实反映最终版图设计的最佳构想。例如，画出大的接触孔，而且采用版图设计所用的彩色。通用的彩色为蓝色代表金属，红色代表多晶，有源区用绿色表示，接触孔用黑色表示，耗尽注入则为黄色，阱区是棕色，以及 p^+ 区为黄色。图 3.34 是对应于图 3.33 中版图的梗图。图 3.35 和图 3.36 给出了两个行家的版图设计，一个是 n MOS，另一个是 CMOS，它们是由 KaThy Tunis 设计的。

3.6 小 结

大生产是器件加工的中心问题。光刻是核心的工艺，其它工艺则是从芯片上去除材料或附加材料。电学限制以及芯片加工机械给版图设计带来了几何图形的限制，设计规则就是从这些限制中拟定出来的。可制造性取决于芯片尺寸以及遵守设计规则的状况。由于不是所有芯片都没有缺陷，所以制造芯片成本的增加速度比芯片面积增长得更快一些。

习 题

- 3.1 试说明如何在 CMOS 工艺中进行场区注入而不另加掩模。
- 3.2 如何在所有 n 阱边缘处的场氧化层下制作 n^+ 保护环。但只能通过调整某些已有掩模而不要另加掩模。
- 3.3 在示例3.2中，假定接触孔与多晶硅对准而不与 n^+ 区对准，试重新导出多晶硅对金属接触的设计规则。
- 3.4 在三模冗余 (triple modular redundancy) 中，制作三组相同的电路，且由主选电路决定输出。假定该主选电路所占面积可以忽略，并设在整个电路面积上 A_{00} 为常数，且泊松统计适用。导出作为 p_0 函数的成品率表达式，其中 p_0 为单个电路能正确工作的几率。画出含有和没有三模冗余时成品率对 p_0 的图形。 p_0 为何值时可用三模冗余一次就把成本赚回来？
- 3.5 已知一个封装在 $40^\circ\text{C}/\text{W}$ 陶瓷管壳内的芯片，在环境温度 23°C 时耗散功率为 1W 。假定器件电

- 随温度的全部变化均起因于迁移率的减少, 试预计芯片工作在环境温度 70°C 时的功耗和结温。
- 3.6 附录所示的 MOS 工艺中, 要求扩散区直线走向时每个 pitch 为 $5\mu\text{m}$, 扩散线怎样才能以 $4\mu\text{m}$ 间距布线而不违反任何设计规则?
 - 3.7 对第一章中所示的电路进行版图设计, 并对有兴趣的部分画出剖面图。
 - 3.8 导出晶体管的硅栅与金属和扩散区的接触孔间最小版图间距的设计规则。假定工艺与示例 3.2 中所述工艺相同。
 - 3.9 设计一对交叉耦合的或非门版图。其工艺中的 V_{DD} 超过了厚场区寄生晶体管的阈值电压。输入和输出电压摆幅都是最低到最高变化的。 $\triangle$
 - 3.10 考虑 CMOS 工艺中的埋孔接触问题。如果我们用 n 型多晶硅通过埋孔接触到 p^+ 材料上, 就会形成 $p-n$ 结二极管。若保证多晶不短接到 n 阱上, 则该二极管就可能是有用的。为了保证不发生短路, 要用什么样的工艺步骤? 另外一种办法是利用 n 阱上的 p^+ 多晶硅和 p 型衬底上的 n⁺ 多晶硅。这样做的问题何在? 为什么多晶硅上的硅化物有助于解决这个问题? $\triangle$
 - 3.11 从图 3.35 和图 3.36 中的版图提取出相应的电路图。
 - 3.12 构思出一种不用高温加工的接触孔自对准工艺方法。 $\triangle\triangle$

参 考 文 献

- [Antoniadis 77] D. A. Antoniadis, S. E. Hansen, R. W. Dutton, and A. G. Gonzalez, "SUPREM 1 — A program for IC process modeling and simulation," Stanford Electronics Laboratories, Stanford, Calif. Tech. Rep. 5019-1, 1977.
- [Antoniadis 79] D. A. Antoniadis and R. W. Dutton, "Models for Computer Simulation of Complete IC Fabrication Process," *IEEE J. Solid-State Circuits* SC-14, 412-422, 1979.
- [Bayless 83] M. Bayless, W. Devanney, W. Waller, and D. Laurent, "An Automated Methodology for Generating Self-Consistent Layout Rules for VLSI Designs," *IEEE International Electron Device Meeting*, 250-254, Washington, D. C., 1983.
- [Blodgett 83] A. J. Blodgett, Jr., "Microelectronic Packaging," *Scientific American* 249, 85-98, 1983.
- [Burger 67] R. M. Burger and R. P. Donovan (eds.), *Fundamentals of Silicon Integrated Device Technology*, Prentice-Hall, Englewood Cliffs, N. J., 1967.
- [Chin 82] K. Y. Chin, R. Fang, J. Lin, J. L. Moll, C. Lage, S. Angelos, and R. Tillman, "The SWAMI—a Defect Free and Near-Zero Bird's-Beak Local Oxidation Process and Its Application in VLSI Technology," *IEEE International Electron Device Meeting*, 224-227, San Francisco, 1982.
- [Chin 83] D. Chin, S-Y Oh, S-M Hu, R. W. Dutton, and J. L. Moll, "Two-Dimensional Oxidation," *IEEE Trans. Electron Devices*, ED-30, 744-749, 1983.
- [Colclaser 80] R. A. Colclaser, *Microelectronics Processing and Device Design*, Wiley, New York, 1980.
- [Conway 80] L. Conway, A. Bell, and M. E. Newell, "MPC79: The Large-Scale Demonstration of a New Way to Create Systems in Silicon," *LAMBDA* 1, (2), 10-19, 1980.
- [Dennard] R. H. Dennard, F. H. Gaensslen, H. N. Yu, V. L. Rideout, E. Bassons, and A. R. LeBlanc, "Design of Ion-Implanted MOSFET's with Very Small Dimensions," *IEEE J. Solid-State Circuits* SC-9, 256-267, 1974. (Scaling theory)
- [Ghandhi 69] S. K. Ghandhi, *The Theory and Practice of Microelectronics*, Wiley, New York, 1969.
- [Gibbons 80] J. F. Gibbons and K. F. Lee, "One-Gate-Wide CMOS Inverter on Laser Recrystallized Polysilicon," *IEEE Trans. Electron Devices Lett.* EDL-1, 117-118, 1980.
- [Gupta 83] D. C. Gupta, (ed.), *Silicon Processing*, Amer. Soc. Testing and Materials (ASTM) Philadelphia, Pa., 1983.
- [Hu 84] S. M. Hu, "On Yield Projection for VLSI and Beyond, I. Analysis of Yield Formulas," *IEEE Electron Devices Newsletter* 69, 4-7, 1984.

- [Hunter 81] W. R. Hunter, T. C. Holloway, R. K. Chatterjee, and A. F. Tasch, Jr., "A New Edge-defined Approach for Submicrometer MOSFET Fabrication," *IEEE Trans. Electron Devices Lett.* EDL-2, 4—6, 1981.
- [IEEE 81] *Proc. IEEE* 69, February, 1981.
- [Jansen 81] W. D. Jansen and D. G. Fairbairn, "The Silicon Foundry, Concepts and Reality," *LAMBDA 2*, (1), 16—28, 1981.
- [Kang 81] S. M. Kang, "A design of CMOS polycells for LSI circuits," *IEEE Trans. Circuits and Systems CAS-28*, 838—843, 1981.
- [Lee 80] Y. K. Lee and J. D. Craig, "Polyimide Coatings for Microelectronic Applications," *Division of Organic Coatings and Plastics Chemistry, ACS Preprints* 43, 451—458, 1980.
- [Liu 83] T. M. Liu and W. G. Oldham, "Channeling Effect of Low-Energy Boron Implant in (100) Silicon," *IEEE Trans. Electron Devices Lett.* EDL-4, 59—62, March, 1983.
- [Liu 84] S. S. Liu, G. E. Atwood, E. Y. So, B. J. Wu, R. W. Leftwich, and K. R. Hasslerjian, "1.5 μ Scaled CMOS Microcomputer Technology," *IEEE International Solid-State Circuits Conf.*, 156—157, San Francisco, 1984.
- [Lyon 81] R. F. Lyon, "Simplified Design Rules for VLSI Layouts," *LAMBDA 2*, (1), 54—59, 1981.
- [Mikkelsen 81] J. M. Mikkelsen, L. A. Hall, A. K. Malhotra, S. D. Secombe, and M. S. Wilson, "A NMOS VLSI Process for Fabrication of a 32-Bit CPU Chip," *IEEE J. Solid-State Circuits* SC-16, 542—547, October, 1981.
- [Muramoto 78] S. Muramoto, T. Hosoya, and S. Matsuo, "A New Self-Aligning Contact Process for MOS LSI," *IEEE International Electron Device Meeting*, 185—188, Washington, D. C., 1978.
- [Oldham 82] W. G. Oldham, "Isolation Technology for Scaled MOS VLSI," *IEEE International Electron Device Meeting*, 218—219, San Francisco, 1982.
- [Parrillo 80] L. C. Parrillo, R. S. Payne, R. E. Davis, G. W. Reutlinger, and R. L. Field, "Twin-Tub CMOS-A technology for VLSI Circuits," *IEEE International Electron Device Meeting*, 752—755, Washington, D. C., 1980.
- [Pashley 77] R. Pashley, W. Owen, K. Kokkonen, R. Jecmen, A. Ebel, C. Ahlquist, and P. Schoen, "A High-Performance 4 K Static RAM Fabricated in an Advanced MOS Technology," *IEEE International Solid-State Circuits Conf.*, 22—23, New York, 1977.
- [Peltzer 83] D. L. Peltzer, "Wafer-Scale Integration: The Limits of VLSI?" *VLSI Design* 4, (5), 43—47, 1983.
- [Price 70] J. E. Price, "A New Look at Yield of Integrated Circuits," *Proc. IEEE*, 1290—1291, 1970.
- [Rea 81] S. N. Rea, "Czochralski Silicon Pull Rate Limits," *J. Cryst. Growth* 54, 267—274, 1981.
- [Rideout 79] V. L. Rideout and V. J. Silvestri, "MOSFET's with Polysilicon Gates Self-Aligned to the Field Isolation and to the Source and Drain Regions," *IEEE Trans. Electron Devices*, ED-26, 1047—1052, 1979. (Self-aligned contacts)
- [Rideout 80] V. L. Rideout, J. J. Walker, A. Cramer, "A One-Device Memory Cell Using a Single Layer of Polysilicon and a Self-Registering Metal-to-Polysilicon Contact," *IBM J. Res. Develop.* 24, 339—347, 1980.
- [Rung 81] R. D. Rung, "Determining IC Layout Rules for Cost Minimization," *IEEE J. Solid-State Circuits* SC-16, 35—43, February, 1981.
- [Saito 82] K. Saito and E. Arai, "Experimental Analysis and New Modeling of MOS LSI Yield Associated with the Number of Elements," *IEEE J. Solid-State Circuits* SC-17, 28—33, 1982. (Some yield characterization numbers)
- [Sakai 81] Y. Sakai, T. Hayashida, N. Hashimoto, O. Mimato, T. Masuhara, K. Nagasawa, T. Yasui, and N. Tanimura, "Advanced Hi-CMOS Device Technology," *IEEE International Electron Device Meeting*, 534—537, Washington, D. C., 1981.
- [Sakamoto 80] M. Sakamoto and K. Hamano, "A New Self-Aligned Contact Technology," *IEEE International Electron Device Meeting*, 136—139, Washington, D. C., 1980.

- [Samuelson 80] G. Samuelson, "Polyimide for Multilevel VLSI," *Division of Organic Coating and Plastics Chemistry, ACS Preprints* 43, 1587—1594, 1980.
- [Smith 81] R. T. Smith, J. D. Chlipala, J. F. M. Brindels, R. G. Nelson, F. H. Fisher, and T. F. Mantz, "Laser Programmable Redundancy and Yield Improvement in a 64K DRAM," *IEEE J. Solid-State Circuits* SC-16, 506—514, 1981.
- [Straub 83] R. R. Straub, "Hydrofluoric Acid Burns of the Hand," *Orthopedics* 6, 978—980, 1983.
- [Sze 83] S. M. Sze, *VLSI Technology*, McGraw-Hill, New York, 1983.
- [Theuwissen 82] A. J. P. Theuwissen, "Linear CCD-Imagers with Polyimide Insulation for Double Level Metallization," *IEEE Trans. Electron Devices Lett.* EDL-3, 308—309, 1982.
- [Wollensen 79] D. L. Wollensen, "C-MOS LSI, comparing second-generation approaches," *Electronics* 52, 116—123, September 13, 1979.
- [Yu 81] K. Yu, R. J. C. Chwang, M. T. Bohr, P. A. Warkentin, S. Stern, and C. N. Berglund, "HMOS-CMOS-A Low-Power High-Performance Technology," *IEEE J. Solid-State Circuits* SC-16, 434—439, 1981.

第四章 逻辑抽象

逻辑抽象在数字范畴和电学范畴之间建立了某种联系。在数字范畴内，用“1”和“0”表示信号。而在电学范畴内，信号本质上是模拟的，含有噪声，并用非理想的转换器来处理。在本章中，我们将分析从电学范畴到数字范畴的变换，考查静态和动态的噪声问题、噪声容限以及最坏情况设计，还将介绍许多功能块部件的定量分析。

4.1 逻辑电平的恢复

有效的逻辑信号是用电压范围表示的。代表有效逻辑电平的电压通过数字网络传播时，我们要求能够恢复这个电压。这就是说，在通过数字网络传播时，我们假定逻辑电压会减小，以致于逻辑信号移动到某一特定范围之外。我们希望把这些信号恢复到有效范围。

我们定义 V_{IL} 为加到恢复逻辑门上的有效输入逻辑低电压， V_{IH} 为加到恢复逻辑门上的有效输入逻辑高电压。对于正逻辑有

$$V_{OH} \geq V_{IH} + NM_H \quad (4.1)$$

$$V_{OL} \leq V_{IL} - NM_L \quad (4.2)$$

其中 V_{OL} 和 V_{OH} 是有效的输出逻辑电压。即使在最坏的温度条件和最坏的加工条件下，低电平噪声容限 NM_L 和高电平噪声容限 NM_H 也必定是正的。

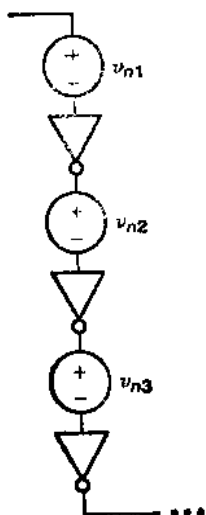


图4.1 加在逻辑门上的噪声电压

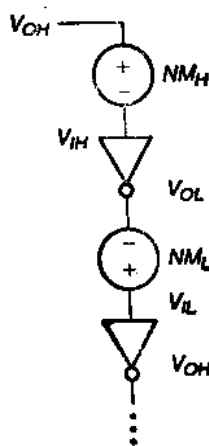


图4.2 反相器链的噪声最坏情况

图4.1是在门之间带有噪声电压源的反相器链。按照噪声容限的定义，如果这些噪声电压低于由噪声容限所定义的极限值，电路就能正确地工作。噪声最坏情况示于图4.2中。

由于在nMOS传输晶体管逻辑电路中存在阈值损失，所以在耗尽型nMOS电路中通常有两个或多个 V_{OL} 和 V_{OH} 值。在我们的nMOS工艺示例中，有三种情况：耗尽型晶体管作为上拉元件，增强型晶体管作为上拉元件，零阈值晶体管作为上拉元件。因而共有三个 V_{OL} 和三个 V_{OH} 值。在CMOS电路中，通常只有一个 V_{OL} 和一个 V_{OH} 值。尽管如此，在CMOS电路中仍可能有多个逻辑电平值，这取决于其中传输晶体管的使用方法。应当注意，在式(4.1)和(4.2)中，我们已经暗暗地假定了并不存在高电平太高以及低电平太低的情况¹⁾。

图4.3为一个典型的反相器传输函数。在坐标上标明了有效逻辑电平的范围。图中的阴影区表示恢复反相器在正常工作情况下传输函数必须处的位置。当输入低于 V_{IL} 时，其输出必定高于 V_{OH} 。当输入处在 V_{IL} 和 V_{IH} 之间时，输出的逻辑值是不定的，可能是某一任意值。当输入高于 V_{IH} 时，其输出必定低于 V_{OL} 。因为 $V_{OH}-V_{OL}$ 一定大于 $V_{IH}-V_{IL}$ ，所以在中间区域各处，反相器必然具有不同的增益。在一个很大的范围内，逻辑电平的赋值完全是任意的，其好处在于有可能利用一个给定的逻辑系列，实现许多种不同的逻辑门。在一个数字逻辑系列中，有效电压值是由某种约定的限制或设计方法决定的。它使一个逻辑系列内的每个门的直流电压传输函数均符合或优于一定的规范，这样，可能有一个或多个门超过这个规范，但不会造成噪声容限和逻辑电平赋值的失效。这表明那些逻辑门可能不是最佳的选择，但这是一个不同的问题，对此我们将在后边讨论。在网络中，当用“1”和“0”来代表某些信号时，噪声容限有助于保证对这些信号所作解释的有效性（当涉及到这些问题时，电路中不能使用时钟方法）。

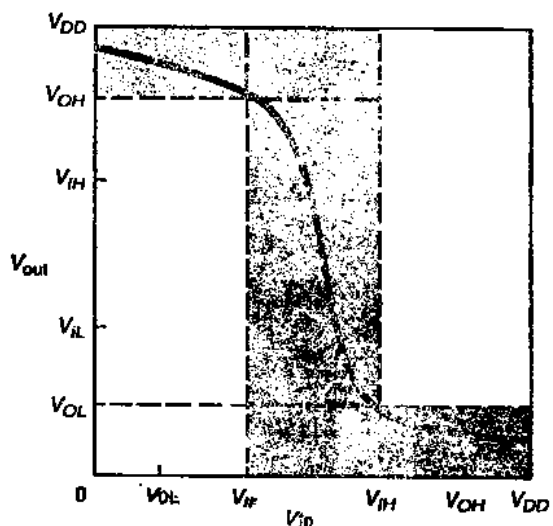


图4.3 标明有效逻辑电压值的反相器传输函数。对于同一逻辑系列，其反相器的传输函数必须处在阴影区域中

由于耗尽型nMOS集成电路中含有两个或多个合法的 V_{OH} 值，因而有时假定若输出高电压只能上升到 $V_{DD}-V_T$ ，就是输出未被恢复；若输出高电压可以上升到 V_{DD} ，就是输出被恢复了。但是这一点并不完全真实。把有效输出电压值的概念与恢复逻辑电平的概念区分开来是很重要的。在一个恢复逻辑门中，输入输出传输函数有一个具有不同增益的区域。图4.4中的头两个电路就是恢复逻辑门，但有效高电平 V_{OH} 的值却为 $V_{DD}-V_T$ 。图4.4(a)是一个增强型晶体管作负载的反相器。图4.4(b)是一个带有控制逻辑

1) 适用的逻辑系列还有许多其它特性。例如，门的输入输出传输特性应能减弱有效逻辑电压的扰动。这就是说，对于所有相应于有效逻辑输入电平的 v_{in} ， $|dv_{out}/dv_{in}|$ 都要小于1，只要这一特性是良好的，其它要求就不严格了。第二个重要特性是门应当单向工作。因此，逻辑信号只能通过恢复门单向传播。很多物理现象可提供增益但却不是单向工作，激光就是一例。除非采用隔离器，否则它的反射就成了问题。然而，除了一些特殊的工艺技术以外，在通常情况下，不必担心这些限制。

辑的与门。图4.4(c)是一个自举传输门，它的输出电压可以达到 V_{DD} ，但它并不是恢复逻辑门电路。图1.37(b)是只有一个恢复输出端的CMOS异或电路。

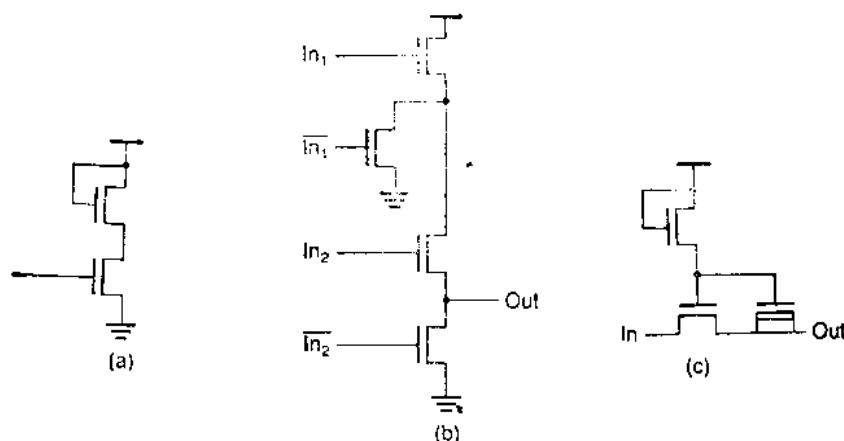


图4.4 逻辑门。(a)具有弱输出高电平的恢复逻辑门；(b)具有弱输出高电平的恢复逻辑门；(c)具有强输出高电平的非恢复逻辑门

为了区别输出电压和恢复逻辑电压这两个概念，我们把输出电压与电源电压接近的程度看作是信号的“强度”。这就是说，一个能上升到 V_{DD} 的输出信号是强输出，而一个只能升到 $V_{DD}-V_T$ 的输出信号就是弱输出。这个考虑与该输出是否被恢复了是不相干的。与此相似，也可以有强低电平信号和弱低电平信号之分。一个弱低电平信号可以仅仅低于300mV，而强低电平信号则必须是地电平。例如，这里的后一种情况相当于输入信号加到控制动态电荷存储的传输管的栅上。由于次开启漏电，晶体管的 v_{GS} 必须远远低于它的开启电压，以便使通过晶体管的电荷慢慢泄漏，直到与p-n结二极管反向漏电流值具有同样的量级。

在我们的逻辑门系列中，所有初级门（诸如反相器、与非门和或非门）都是反相的。定义任一输入端和输出端之间的静态传输函数如下：

$$V_{out} = H(V_{in}) \quad (4.3)$$

将式(4.1)和(4.2)用于反相器，可有

$$H(H(V_{OH} - NM_H) + NM_L) \geq V_{OH} \quad (4.4)$$

$$H(H(V_{OL} + NM_L) - NM_H) \leq V_{OL} \quad (4.5)$$

图4.2说明了这一点。可以把式(4.4)和式(4.5)推广到处理或非门和与非门。对于一个与非门，如果所有输入都大于或等于 V_{IH} ，则输出必然小于或等于 V_{OL} 。如果任一输入小于或等于 V_{IL} ，则输出必为 V_{OH} 。对此，可以用多维几何作类似于图4.3的解释。图4.5对与非门和或非门网络的噪声最坏情况作了说明。

在给出基本的噪声容限概念之后，再看看它的某些派生的特性。用使噪声容限之和最大的方法来选定四个逻辑电压的指定值，通过观察由同样反相器组成的长链，可以证明这个目标函数。由于任一级向高电平转换都会引起下一级向低电平转换，所以，使高电平噪声容限与低电平噪声容限之和最大，就能使整个系统的噪声容限最大。我们将

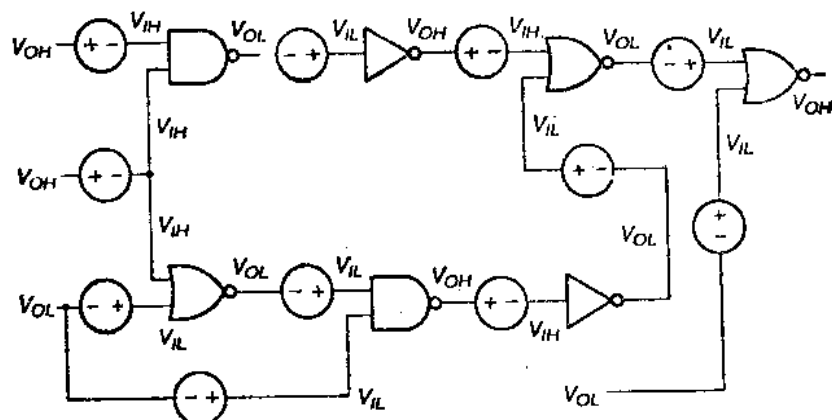


图4.5 逻辑门所汇集的噪声最坏情况。噪声低于这些极限时，电路可保证正确工作

讨论易于实现此点的情况和不能实现此点的情况。对于反相器，噪声容限之和可由下式给出：

$$NM_L + NM_H = V_{IL} - V_{IH} + H(V_{IL}) - H(V_{IH}) \quad (4.6)$$

求解 V_{IL} 和 V_{IH} 的最佳值，我们得到

$$-1 = \frac{dH(V_{IL})}{dV_{IL}} \quad (4.7)$$

和

$$-1 = \frac{dH(V_{IH})}{dV_{IH}} \quad (4.8)$$

注意，式(4.7)和(4.8)是在一定的受约束的假设下导出的，它们并不是噪声容限基本定义的一部分。把这个结果转换到扇入大于1的逻辑门，式(4.7)和(4.8)就表示，若或非门的所有输入端都连在一起且上升到 V_{IL} 值，则这种输入端的电压的扰动将在输出端引起一个大小相等符号相反的扰动。这就意味着当所有输入为低电平时，任一独立的逻辑输入端对输出的小信号增益幅度必小于1。比方说，若一个或非门有 M 个同等的输入端，在最佳 V_{IL} 的情况下，其微分增益为 $-M^{-1}$ 。因为，假如几个输入端的小信号增益都是1，则该门电路将如同一个噪声加法器。这一点很有意义。应当注意，这里导出的最佳逻辑电压值，是假定了这些输入端连在一起，并且，系统中每个逻辑门的传输函数都简化为传输函数为 H 的单个反相器。后边我们将用CMOS列举一个这方面的例子。

噪声容限的定义是：整个系统在每条连线上受到各种方式的噪声干扰，当这些干扰达到噪声容限的极限时，该系统仍旧能够正常工作。这就是说，如果系统中的噪声低于噪声容限，则该系统能够保证它的功能。我们可以回答与此相反的问题：在什么噪声水平时系统就不能保证它的功能了。假定系统中只有一个噪声脉冲加在一条连线上，若该噪声没有使受影响逻辑门的输出跨过 V_{inv} 值，则干扰过后系统最后会恢复原状（为了简化，再假定所有逻辑门都具有理想的传输特性）。若噪声变化使该输出跨过了 V_{inv} 值，就会产生误动作。此处 V_{inv} 代表输出等于输入时的电压值。即

$$V_{inv} \equiv H(V_{inv}) \quad (4.9)$$

V_{inv} 和 V_{OL} 或 V_{OH} 之间的电压差，可作为电路对于噪声忍耐程度的一个不太全面的衡量。噪

声耐度的衡量可由下式给出:

$$NR_L \equiv V_{inv} - V_{OL} \quad (4.10)$$

$$NR_H \equiv V_{OH} - V_{inv} \quad (4.11)$$

在数字系统中,电压并不是与噪声有关的唯一参数。阻抗虽然定义不太明确,但却是另一个与噪声有关的重要因素。所有类型的电容耦合噪声,在高阻抗节点上会比在低阻抗节点上产生更大的噪声电压。由于这些节点都是绝缘的,所以那些存储动态电荷的节点,是对噪声最敏感的节点,各预充节点都属于这一类。在nMOS电路中,把节点预充为高电平不会出现故障。同时 NM_H 远远大于 NM_L ,这就是nMOS预充电路能承受高噪声电压的原因。对CMOS电路设计,节点可以预充为高电平也可以预充为低电平,但是由于被预充的节点后边通常总是跟着反相器,所以在各预充节点上可以恢复高的噪声电压。

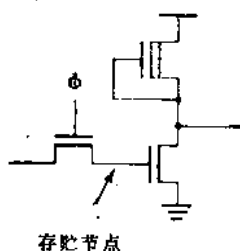


图4.6 对耦合噪声敏感的动态节点

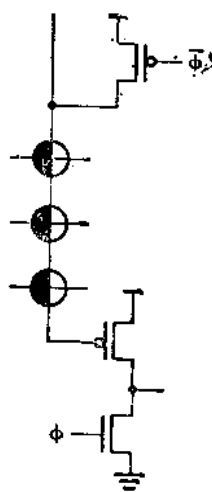


图4.7 对噪声非常敏感的带有预充的CMOS或门。如果高阻抗预充总线的电位变化高达 $|V_{TPO}|$,则输出将会翻转

在这些例子中,噪声容限和阻抗的大小是一致的。但在另一些情况下有可能是不一致的。所以,在版图设计中要非常小心,以便使噪声减至最小。nMOS移位寄存器单元中的动态存储节点就是这种电路的一个例子,如图4.6所示,该存储节点的电容值很低,又加上它是绝缘的,所以易受耦合干扰,最好不要在这个节点的上边跨过金属信号线。在此,还必须考虑传输晶体管栅源电容造成的噪声耦合。在nMOS和CMOS电路中都存在一种特别敏感的结构,图4.7为CMOS电路的一个例子,在这个电路中,如果高阻抗节点上的电位变动一个阈值电压,输出就会翻转。如果用一个完全互补的反相器代替预充电路,则意味着要想使输出翻转,高阻抗节点的电位就必须一直变到 V_{inv} 附近。

nMOS电路中的噪声容限

在MOS集成电路中,为便于对芯片上参数的跟踪,起码要把一个逻辑电平规定成是与晶体管阈值电压相关联的量,而不是表示成某个固定的电压。在nMOS电路中,我们

把 V_{IL} 规定成与 V_{TE0} ($V_{BS}=0$ 时的 V_{TE} 值) 相关的量¹⁾。

在nMOS电路中, 噪声容限问题的关键在于或非门。按照定义, 如果加到或非门上的任一输入电压至少为 V_{IH} , 则或非门的输出电压必定小于或等于 V_{OL} 。这就使每个下拉管的形状因子与上拉管的形状因子的比例大于某个常数, 这个比例通常称作 β 比。我们定义 β 为:

$$\beta \equiv \frac{S_{pd}}{S_{pu}} \quad (4.12)$$

其中 S_{pd} 为下拉管的形状因子, S_{pu} 为上拉管的形状因子。 V_{OL} 的这种限制和或非门(其特例反相器也包括在内)的扇入数无关。若或非门的所有输入都小于或等于 V_{IL} , 它的输出电压必定大于或等于 V_{OH} 。这是另一个限制。只有所有下拉管的关断都是充分的, β 比的限制才和扇入数无关。这就意味着具有关系式²⁾

$$V_{IL} \leq V_{TE0} - \sigma V_{OH} \quad (4.13)$$

在式(4.13)中, 包含了因漏感应势垒降低而引起的阈电压的下降。对于耗尽型负载的nMOS电路, 我们有 $V_{OH}=V_{DD}$, 当电源电压为5V时, 典型的低电平噪声容限值 NM_L 为100—200mV, 高电平噪声容限值 NM_H 为800—1000mV。

可以把式(4.13)和导出的最佳 V_{IL} 值联系起来。如果取斜率为-1的点作为或非门的扇入数上升到无限大时的极限, 就可以发现, 这些极限与上边讨论过的 V_{IL} 和 V_{OL} 值相吻合。把 V_{IL} 取为增强型下拉管的阈电压后, 我们就同时选取了 V_{IL} 作为对应于不同扇入数的输入低电平最佳值中的最小值。对于预充门, 噪声容限为 $V_{TE0} - \sigma V_{OH} - V_{OL}$, 它相应于 $\beta = \infty$ 时的有比情况。图4.8给出了nMOS耗尽型负载反相器的传输函数, 该图以下拉器件的栅宽 W 作为参变数。

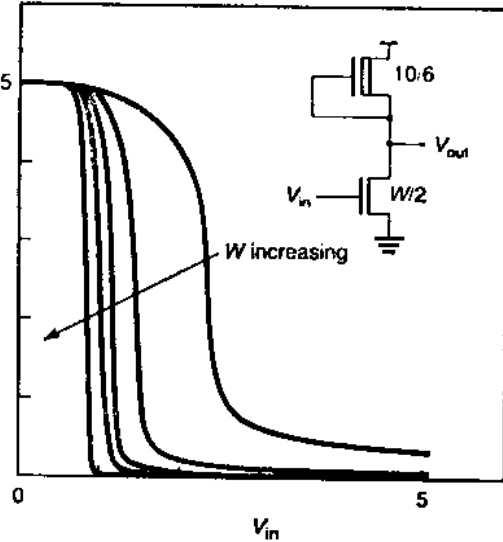


图4.8 耗尽负载nMOS反相器的传输特性。图中给出了下拉管栅宽 W 上升时特性的变化

- 1) 在第二章中, 我们曾把衬底作为所有电压量的参考点。从器件物理角度分析晶体管, 这样作是恰当的。然而在考虑电路问题时, 则希望以地电位(信号地)作为电压量的参考, 而不再用衬底作为参考点。在很多nMOS电路技术中, 把衬底偏置为相对于信号地为负的电压。所以要变换坐标使衬底为负电压 V_{BB} , 各个电压均以信号地作为参考。阈电压取自相应的衬底偏置电压。对于 $V_{BB}=0$, 有 $V_{TX}=V_{TE0}$, 而对于 $V_{BB}<0$, 则有

$$V_{TX} = V_{TE0} + \gamma (\sqrt{-V_{BB} + 2\phi_{FP}} - \sqrt{2\phi_{FP}})$$

V_{TE0} 为 $V_{BS}=0$ 时(在此, $V_{BS}=V_{BB}$)的阈电压值, 它是几何结构的函数。对于CMOS电路, 通常 $V_{BB}=0$, 我们用 V_{TN0} 代表当 $V_{BS}=0$ 时nMOS增强型器件的阈电压, 用 V_{TP0} 代表当 $V_{BS}=0$ 时pMOS增强型器件的阈电压。

- 2) 在大的PLA中, 下拉管是非常小的, 但个数却很多, 因而流过那些即使被认为是关断的器件的泄漏电流也会成为问题。

CMOS电路中的噪声容限

对于CMOS电路,规定有效逻辑电平的最直接方法是让 $V_{OH}=V_{DD}$ 和 $V_{OL}=0$ 。这样一来, n 沟和 p 沟增强型器件的阈值电压就分别是低电平和高电平噪声容限。虽然常常是这样处理的,但是存在噪声容限之和不是很大的问题。这是因为当阈值电压较低时,为了使输出电平达到 V_{DD} 或地电平,所需 V_{in} 和 V_{in} 的值是接近于地电平或 V_{DD} 的。

为了获得好的性能,传统CMOS电路的扇入数往往比较小。在传统的CMOS电路中,每个 n 型器件都有一个 p 型器件与之对应。在 nMOS 电路中,为了确定 V_{in} 值,我们取极限扇入数 $M \rightarrow \infty$ 。而对于CMOS电路并不需要这样作。例如,为了确定 V_{in} 和 V_{in} 值,可以选取与非门和或非门($M=1, 2, 3$)的最佳 V_{in} 和 V_{in} 值分别为最小和最大。

传统CMOS逻辑门的“推-拉”特性导致高的微分增益,并因而具有十分理想的传输函数。所以传统CMOS电路具有非常好的噪声特性,但当使用预充电电路或其它先进的CMOS电路技术时,不能认为CMOS电路一定具有好的噪声性能。

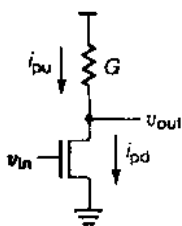
4.2 反相器

MOS反相器含有一对MOS晶体管,它们连接在一起,能实现由式(4.4)和式(4.5)所规定的传输函数。MOS反相器有许多种可能的结构,我们将分析其中的四种,并且对一些简单的MOS电路推导出一系列公式,许多结果要在后边的章节中使用。除作为实用的参考之外,这些结果的推导还有另外两个目的:首先,通过它们说明一些电路分析技巧;其次,它们常常指出某些意外的参量灵敏度和依赖关系。MOS反相器有两种基本类型。第一类反相器中的两个器件都是有源元件,在第二类电路中,有一个器件为无源负载。下面,我们先看看具有理想电阻负载的nMOS反相器。

电阻负载nMOS反相器

电阻负载nMOS反相器如图4.9所示,这是一个最简单的MOS电路。对于具有高体效应的耗尽型器件,或栅电压远高于 V_{DD} 的增强型器件,都能以这种结构作为准确的模型。当然,它也是以多晶硅电阻为负载的反相器的极好模型。流过上拉电阻的电流由下式给出:

$$i_{pu} = (V_{DD} - v_{out}) G \quad (4.14)$$



流过增强型下拉管的电流为

$$i_{pd} = \begin{cases} 0 & v_{in} \leq V_{T0} \\ \frac{K'S}{1+\delta} (v_{in} - V_{T0})^2 & 0 < v_{in} - V_{T0} < (1+\delta)v_{out} \\ K'S(2(v_{in} - V_{T0})v_{out} - (1+\delta)v_{out}^2) & v_{in} - V_{T0} \geq (1+\delta)v_{out} \end{cases} \quad (4.15)$$

图4.9 电阻负载反相器

令上边两个电流相等,可以得到该反相器的稳态传输函数。

$$V_{out} = \begin{cases} V_{DD} & V_{in} \leq V_{TE0} \\ V_{DD} - \frac{1}{B} \left(\frac{V_{in} - V_{TE0}}{1 + \delta} \right)^2 & 0 < V_{in} - V_{TE0} < (1 + \delta)V_{out} \\ \frac{V_{in} - V_{TE0}}{1 + \delta} + \frac{B}{2} - \left(\left(\frac{V_{in} - V_{TE0}}{1 + \delta} + \frac{B}{2} \right)^2 - BV_{DD} \right)^{1/2} & V_{in} - V_{TE0} > (1 + \delta)V_{out} \end{cases} \quad (4.16)$$

其中

$$B = \frac{G}{K'S(1 + \delta)} \quad (4.17)$$

反转电压 V_{inv} 定义为 $V_{in} = V_{out}$ 的电压点。这时晶体管处于饱和状态。求解上式则得

$$V_{inv} = V_{TE0} - \frac{B(1 + \delta)^2}{2} + \frac{(1 + \delta)}{2} \sqrt{\frac{B^2(1 + \delta)^2}{4} + B(V_{DD} - V_{TE0})} \quad (4.18)$$

小信号增益最大的点也是处在饱和区。可用下述两种方法求解小信号增益：

第一种方法：将 V_{out} 对 V_{in} 微分，在 $V_{in} - V_{TE0} \leq (1 + \delta)V_{out}$ 条件下可有

$$\frac{\partial V_{out}}{\partial V_{in}} = a_0 = - \frac{2K'S}{(1 + \delta)G} (V_{in} - V_{TE0}) = - 2 \sqrt{\frac{K'S}{(1 + \delta)G}} \sqrt{V_{DD} - V_{out}} \quad (4.19)$$

由上式可以看出，微分增益 a_0 随 K' 和 S 上升而上升，并且随 δ 和 G 上升而下降。在 V_{inv} 点增益的绝对值达最大。

$$a_{0max} = 1 - \sqrt{\frac{1}{4} + \frac{V_{DD} - V_{TE0}}{(1 + \delta)^2 B}} = 1 - \sqrt{\frac{1}{4} + \frac{(V_{DD} - V_{TE0})K'S}{(1 + \delta)G}} \quad (4.20)$$

对于恢复逻辑电平的反相器， $|a_0|$ 必须大于1。由式(4.20)可以看到， $|a_{0max}|$ 仅取决各主要参数的平方根值。这是MOS电路难以获得高微分增益的原因之一。

第二种方法：先要导出晶体管的小信号模型。我们定义晶体管的跨导 g_m 为：

$$g_m = \left. \frac{\partial i_{DS}}{\partial V_{GS}} \right|_{V_{DS}} \quad (4.21)$$

可有

$$g_m = \begin{cases} 0 & V_{GS} \leq V_{TE0} \\ \frac{2K'S}{1 + \delta} (V_{GS} - V_{TE0}) & 0 < V_{GS} - V_{TE0} < (1 + \delta)V_{DS} \\ 2K'SV_{DS} & V_{GS} - V_{TE0} \geq (1 + \delta)V_{DS} \end{cases} \quad (4.22)$$

我们定义输出电导为

$$g_D = \left. \frac{\partial i_{DS}}{\partial V_{DS}} \right|_{V_{GS}} \quad (4.23)$$

求解¹⁾可得

1) 实际上饱和区的输出电导 g_D 并不是零。由于漏感应势垒降低、沟道长度的调制以及源和漏区串联电阻的影响，使得 g_D 并非理想值。

$$g_D = \begin{cases} 0 & V_{GS} - V_{TH0} \leq (1 + \delta)V_{DS} \\ 2K'S[V_{GS} - V_{TH0} - (1 + \delta)V_{DS}] & V_{GS} - V_{TH0} \geq (1 + \delta)V_{DS} \end{cases} \quad (4.24)$$

利用基本电路分析技术，得到反相器增益由下式给出：

$$a_0 = -\frac{g_m}{g_D + G} \quad (4.25)$$

在进行适当的代换后上式将化为式 (4.19)。

对稳态传输函数进行微分，可以得到所有输入情况下的小信号增益：

$$a_0 = \begin{cases} 0 & V_{in} < V_{TH0} \\ -\frac{2(V_{in} - V_{TH0})}{(1 + \delta)^2 B} & 0 < V_{in} - V_{TH0} < (1 + \delta)V_{out} \\ \frac{1}{1 + \delta} \left\{ 1 - \left[\left(\frac{V_{in} - V_{TH0}}{1 + \delta} + \frac{B}{2} \right)^2 - BV_{DD} \right]^{-1/2} \left(\frac{V_{in} - V_{TH0}}{1 + \delta} + \frac{B}{2} \right) \right\} & V_{in} - V_{TH0} > (1 + \delta)V_{out} \end{cases} \quad (4.26)$$

反相器传输函数有“0”点、“1”点和两个斜率为-1的点。如果反相器是恢复逻辑门，则存在两个解，它们可由下式求出：

$$V_{in} = \frac{(1 + \delta)^2 B}{2} + V_{TH0} \quad (4.27)$$

$$V_{in} = V_{TH0} + (1 + \delta) \left\{ \left[\frac{(2 + \delta)^2 BV_{DD}}{\delta^2 + 4\delta + 3} \right]^{1/2} - \frac{B}{2} \right\} \quad (4.28)$$

可以把电阻负载反相器的一些参数和在 4.1 节中导出的噪声容限联系起来。利用定义 $V_{IL} \equiv V_{TH0}$ 和 $V_{OH} \equiv V_{DD}$ ，就有

$$V_{OL} = \frac{V_{IH} - V_{TH0}}{1 + \delta} + \frac{B}{2} - \left[\left(\frac{V_{IH} - V_{TH0}}{1 + \delta} + \frac{B}{2} \right)^2 - BV_{DD} \right]^{1/2} \quad (4.29)$$

其中 V_{IH} 由式 (4.1) 给出。实际情况与此有些细微的区别。我们记得，由于漏感应势垒降低， V_{TH0} 实际上和 V_{DD} 有关， V_{TH0} 也是晶体管沟道宽度和长度的函数。噪声容限这个概念应用到一个逻辑门上时，这个逻辑门驱动的不是它本身而是另外一个逻辑门。如果这两个逻辑门中管子的沟道宽度和长度是一样的， V_{TH0} 对这些参数的依赖关系就可以不予考虑（但是，如果它们是不同的，就应当把变化关系考虑进去）。观察刚刚写出的公式，可以看出，如果式 (4.16) 是与所讨论的门有关的，则 $V_{IL} = V_{TH0}$ 就与下一级门有关。因为在下级门输入为 V_{IL} 时它的输出为“高”，可以更精确的写成 $V_{IL} = V_{TH0} - \sigma V_{DD}$ ，该式与式 (4.13) 是一致的。这样一个细微的差别对于电路设计可以产生很大影响。

电阻负载反相器的瞬态特性是不对称的。为了简化，假定它的输入是一个从 0 到 V_{DD} 或从 V_{DD} 到 0 的阶跃电压。大多数情况下，逻辑门的瞬态特性取决于非本征电容而不是本征电容¹⁾。下面分析反相器驱动一个线性负载电容 C_L 的情况。若 $t = 0$ 时输入信号

1) 本征电容是那些和晶体管直接相关联的电容，就是式 (2.119) 和 (2.120) 中的电容。

由高转变为低，则输出的上升瞬态特性取决于负载电导 G 。可有

$$v_{out}(t) = (V_{DD} - V_{OL})(1 - e^{-tG/C_k}) + V_{OL} \quad (4.30)$$

如果输入信号由 0 转变为 V_{DD} ，则情况更加复杂。达到某一特定电压 v_{out} 的时间由下式给出：

$$t = \int_{V_{DD}}^{V_{out}} \frac{C}{i_{pu} - i_{pd}} dv_{out} \quad (4.31)$$

其中 i_{pu} 和 i_{pd} 由式 (4.14) 和 (4.15) 给出，此时 $v_{in} = V_{DD}$ ，(参看附录 B) $v_{out}(t)$ 的波形在一开始是线性的。其特性与电流源驱动电容时电容上的电压特性相同。接近末尾处 $v_{out}(t)$ 的波形呈指数形。

实际上，对应于输入阶跃波形的下降沿，反相器的输出电压会下降到稍低于地电平，这是由于密勒 (Miller) 电容 C_{GD} 所起的电荷泵作用所致。密勒电容对电路的瞬态响应有很大的影响，但现在数字逻辑电路，因而它的整个影响是有限的，这时在 C_{GD} 上电荷的变化极限值为 $2C_{GD}V_{DD}$ 。对于由非零阻抗源驱动的逻辑门，密勒电容可引起电压瞬态波形的凹陷。

在极端情况下，电阻负载反相器的有效输入电容可以从 $\sim 3\%WLC_{ox}$ 变到 $\sim 130\%WLC_{ox}$ ，这取决于输入波形的变化速度和变化方向。这里的 W 和 L 分别表示反相器下拉管的栅宽和栅长。对于由低变到高的输入电压，下拉 MOS 晶体管的本征电容开始时接近于 0，在输入达到阈值电压时，电容值很快升到大约 $\frac{2}{3}WLC_{ox}$ 。随着反相器输出电压的下降，密勒电容将其变化耦合到输入电容上。在瞬态过程基本结束以后，反相器的输入电容为 WLC_{ox} 。反相器的有效输入电容可能从小于 $\frac{2}{3}WLC_{ox}$ (对于变化非常快的输入) 变到由密勒电容决定的很大的数值，这依赖于输入与输出瞬变时间之间的相互关系。对于变化较慢的输入而言，情况类同。对于由高到低输入下降的瞬变，电容数值的变化次序则不相同。所有这些表明要想一眼看出通过一个器件链路的延时似乎是设计师力所不及的。通常可以采用简单的延时模型去估计延迟时间，其误差约为 30%。但是，这里所讨论的一些效应，确实限制了这些简单估计的精度。

器件尺寸越来越小时，前边所讨论的许多效应变得越来越严重。既然如此，为什么我们还要力求推向短沟道器件呢？可以通过分析一对有比逻辑门找到这个问题的答案。逻辑门中的功耗主要决定于上拉管。现在我们来考查一个尺寸给定的上拉管和栅长变化的下拉管组成的电路，看看它的延时会发生什么变化。下拉管的形状因子 S_{pd} 必须保持为常数，以便使该逻辑门能够满足噪声容限的限制。逻辑门的输入电容正比于 $W_{pd}L_{pd}$ ，因而，由固定尺寸的上拉管驱动逻辑门输入电容时，其延时和功耗-延迟乘积也正比于 $W_{pd}L_{pd}$ (功耗是常数)。但是由于 S_{pd} 是常数，而 $W_{pd} = S_{pd}L_{pd}$ ，因此，延时和功耗-延迟乘积都正比于 L_{pd}^2 。这就是缩小 L_{pd} 的强大动力。在 CMOS 电路技术中，对于一个给定栅宽的晶体管，也可以导出电路的延时依赖于 L^2 的关系。

耗尽负载 nMOS 反相器

在有比 nMOS 电路中，与下拉管相比，电路的负载 (它连接在输出端和 V_{DD} 之间) 既读不是更多地，也是同样地对逻辑门的性能起着决定作用。假定负载是一个两

端无滞后效应的电阻性元件，则它的电流-电压特性存在一些限制。若已知上拉元件是无源的，当输出达到 V_{DD} 时流过负载的电流应该为 0。此外，当输入处在 V_{TH} 时，输出端应该是 V_{OL} ，此时流过上拉元件的电流必等于流过下拉管的电流。图 4.10 为四种不同上拉元件的 $I-V$ 特性曲线。它们都可以采用同样的下拉管。当逻辑门的输出为 V_{OL} 时，它们全都流过同样的电流。用增强型晶体管作上拉元件时，输出电平不能上拉到如同其它三种上拉元件达到的那么高。除此之外，从静态的观点来看，别的都是等价的。而从动态特性来看，则有很大不同。如果输出处在 V_{OL} 时关断下拉管，输出将上升，输出上升到电压 V_{finish} 的时间由下式给出：

$$\Delta T = C_{LOAD} \int_{V_{DD}-V_{OL}}^{V_{DD}-V_{finish}} I_{pu}^{-1} dV \quad (4.32)$$

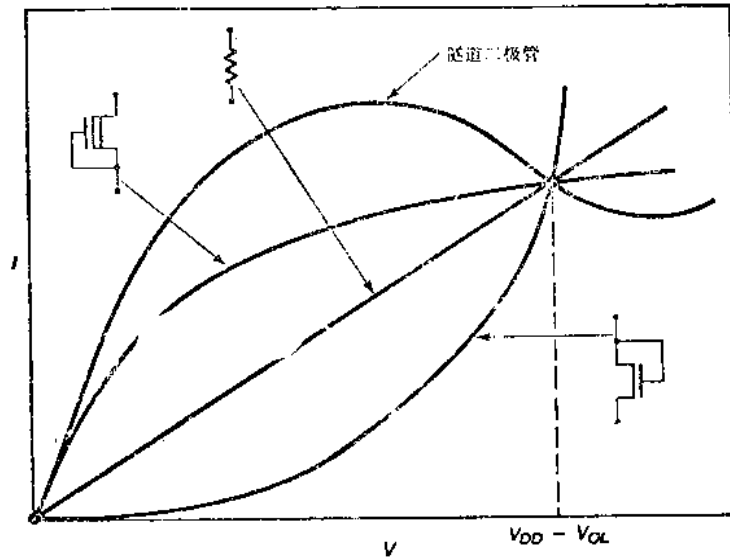


图 4.10 各种负载元件 $I-V$ 特性的比较

可以采用 $V_{finish} = V_{TH}$ 或者取 V_{finish} 略低于 V_{TH} 值。式 (4.32) 的具体内容随上拉元件而异。增强型管作负载是其中最差的，线性电阻作负载次之，耗尽型管作负载是常用负载中的最佳者，而且体效应越小其特性越好。从理论观点来看，隧道二极管作负载应该是最理想的，但这是不实际的。隧道二极管具有负的微分电阻，所以可能很不稳定，除非用下拉管的正电阻把它补偿掉。我们之所以把隧道二极管负载包括在讨论之中，只是为了指出在这个领域内仍旧有新概念的发展余地。在 nMOS 电路中，功率耗散主要发生在逻辑门输出为低电平的时候，通常动态开关功耗小于准直流功耗值的 5%。如果忽略动态功耗，我们曾经说过，每一种上拉元件都将抽取大约相同瓦数的功率，因此式 (4.32) 所给出的延时就正比于功耗-延迟积。与增强型负载或电阻负载逻辑系列相比，耗尽负载逻辑系列的功耗-延迟积更佳。

现代的耗尽型负载 nMOS 管具有很低的体效应，这是因为衬底掺杂较轻和栅氧化层很薄的缘故。因而在这类器件中通常可以忽略 δ 项。上拉管中的电流由下式给出：

$$i_{pu} = \begin{cases} K'_D S_D (-2V_{TD} - V_{DD} + v_{out})(V_{DD} - v_{out}) & V_{DD} - v_{out} \leq V_{TD} \\ K'_D S_D V_{TD}^2 & V_{DD} - v_{out} \geq V_{TD} \end{cases} \quad (4.33)$$

通常 V_{TD0} 在 -2.5V 左右。由于体效应, V_{TD} 是 v_{out} 的函数。在图 4.8 中画出了典型的 $v_{out}-v_{in}$ 传输函数。有两个工作区要特别注意。当两个器件都处在饱和状态时, 微分电导最大。忽略掉 δ 和 σ 项, 饱和下拉管的小信号特性为

$$-\frac{\partial i_{pd}}{\partial v_{in}} = 2\sqrt{K'_S S_E I_{pd}} \quad (4.34)$$

对于上拉管可有

$$\frac{\partial i_{pu}}{\partial v_{out}} = -2b\sqrt{K'_D S_D I_{pu}} \quad (4.35)$$

其中

$$b = \frac{\partial V_{TD}}{\partial v_{out}} \quad (4.36)$$

计算 b 可得¹⁾

$$b = \frac{\gamma}{2\sqrt{V_{SB} + 2\phi_{FP}}} \quad (4.37)$$

根据前述负载特性, 为了达到最佳功耗-延迟积, b 应当尽可能地低些。这就意味着我们希望低的 γ 和高的 V_{SB} 值。这就是在高性能 nMOS 电路技术中采用负 V_{SB} 偏置的原因之一。对于低 γ 值的工艺, 还要保持低 $|V_{TD}|$ 值, 以便使上拉管尽可能长时间地处在饱和状态, 这样有利于在最大的上拉电压范围内使 i_{pu} 保持常数, 而且还能保持低的上拉管电容。应当注意, 在我们的模型中 $b = \delta$, 这是我们对特殊模型进行假设的人为结果, 不能把它看作不变的真理。耗尽型负载反相器的小信号增益可通过使 i_{pu} 和 i_{pd} 相等而得到。我们有

$$a_0 = \frac{\partial v_{out}}{\partial v_{in}} = -\frac{1}{b} \sqrt{\frac{K'_E S_E}{K'_D S_D}} \approx -\frac{\sqrt{\beta}}{b} \quad (4.38)$$

对于很高增益的短沟道器件, 漏感应势垒降低和沟道长度调制可成为它的限制因素。例如, 即使负载是一个电流为 I_{pu} 的理想电流源, 小信号增益的大小也将限制为 σ^{-1} 。可有

$$I_{pu} = K'_E S_E (v_{in} - V_{TE0})^2 \quad (4.39)$$

对 v_{in} 微分, 可得

$$\begin{aligned} 0 &= 2K'_E S_E (v_{in} - V_{TE0}) \frac{\partial}{\partial v_{in}} (v_{in} - V_{TE0}) \\ &= 1 - \frac{\partial v_{out}}{\partial v_{in}} \frac{\partial V_{TE0}}{\partial v_{out}} = 1 + a_0 \sigma \end{aligned} \quad (4.40)$$

在此我们认为

$$a_0 = \frac{\partial v_{out}}{\partial v_{in}} \quad (4.41)$$

$$\sigma = -\frac{\partial V_{TE0}}{\partial v_{out}} \quad (4.42)$$

1) 应当指出, 大多数 MOS 器件的模型参数是通过使大信号参数与测量曲线拟合而得到的。小信号参数通常指的是微分, 因而这样得到的参数值可能与它们的真实数值相差很远。

求解式 (4.40) 得到

$$a_0 = -\frac{1}{\sigma} \quad (4.43)$$

当耗尽型负载反相器的增益受这个机制限制时, 可以增加下拉管的实际沟道长度, 以便使它脱离短沟道状态。

输出接近 V_{OL} 值, 是它的第二个重要状态。假定阈值电压适中, 并且上拉管处在饱和状态。使流过上拉管和流过下拉管的电流相等, 则得

$$K'_D S_D V_{TD}^2 = K'_E S_E [2 (v_{in} - V_{TE0}) v_{out} - v_{out}^2] \quad (4.44)$$

这里假定了下拉管处在非饱和状态。在此状态下, $V_{TD} \approx V_{DD}$ 。当 $v_{in} = V_{IH}$, $v_{out} = V_{OL}$ 时, 式 (4.44) 限定了 β 比, 可有

$$\beta \geq \frac{K'_D}{K'_E} \frac{V_{TD}^2}{[2 (V_{IH} - V_{TE0}) V_{OL} - V_{OL}^2]} \quad (4.45)$$

耗尽型负载反相器的动态特性是相当复杂的。动态特性开始的40%是重要的, 我们集中分析这一部分的输出波形。假定有一个阶跃输入, 增强型晶体管决定了下拉管的瞬态特性。如果输入从 0 变到 V_{DD} , 则下拉电流初始值为

$$i_{p0} = K'_E S_E (V_{DD} - V_{TE0})^2 \quad (4.46)$$

在输出端, 初始的电压摆率由下式给出:

$$\frac{dv_{out}}{dt} \approx \frac{K'_E S_E (V_{DD} - V_{TE0})^2}{C_{LOAD}} \quad (4.47)$$

在上拉瞬变期间, 下拉管是关断的。上拉管开始处在饱和状态。可有

$$I_{p0} = K'_D S_D V_{TD}^2 \quad (4.48)$$

或

$$\frac{dv_{out}}{dt} \approx \frac{K'_D S_D V_{TD}^2}{C_{LOAD}} \quad (4.49)$$

选择 $|V_{TD0}|$ 时要进行折衷。如果缩小 $|V_{TD0}|$ 值, 上拉管的版图将很紧凑。但是阈值电压的不确定性与阈值的大小无关。当 $|V_{TD0}|$ 大时, 延时将更加受到控制。求解上拉瞬变对于下拉瞬变的比值 β_r , 可有

$$\beta_r = \frac{\text{上升电压摆率}}{\text{下降电压摆率}} = \beta \frac{K'_E}{K'_D} \frac{(V_{DD} - V_{TE0})^2}{V_{TD}^2} \quad (4.50)$$

用式 (4.45) 中的 β 代入, 式 (4.50) 化简为

$$\beta_r = \frac{(V_{DD} - V_{TE0})^2}{[2 (V_{IH} - V_{TE0}) V_{OL} - V_{OL}^2]} \quad (4.51)$$

从一级近似看, β_r 和 V_{TD} 无关而随 V_{DD} 上升。对于合理的阈值电压, 选择 $\beta_r \approx 2\beta$ 。这个趋势主要受三个因素的影响: 首先, 晶体管并非处在饱和状态; 其次, 输入不是阶跃函数 (它使下降时间的变长大于上升时间的变长); 第三, 通常开关点 V_{inv} 离地电平比离 V_{DD} 更近些 (相对于上升时间而言, 它延长了有效的下降时间。)

图 1.28 是两个典型 nMOS 反相器的版图。其中第一个反相器的负载包括有固有的负载电容、连线电容和下一级的输入电容。固有的负载电容有几个成分, 计有上拉管沟道对衬底的电容、隐埋接触的电容、延伸多晶硅的电容 (正比于上拉管的栅长)、扩散电容 (正比于下拉管的栅宽)、上拉管的栅漏电容 (由于上拉管处在饱和状态, 所以这

个电容主要是交叠电容)等等。当器件比较小的时候,在上述这些自身固有之中,隐埋接触的电容往往是主要的。连线电容曾在2.4节中讨论过。

对于具有非常缓慢斜坡的输入信号,反相器的瞬态响应决定于它的直流传输函数。逻辑电压的转变时间定义为电压从一个有效值变到另一个有效值所占用的时间。逻辑门输出端的转变时间大约等于输入转变时间除以该逻辑门的“增益”。设想有一个由相同反相器组成的长链,如果从第一级开始输入的信号就具有极长的转变时间,那么,当信号通过该链传播时,转变就会越来越快直至达到稳定状态。在稳态情况下,转变时间受限于负载电容与直流传输功能两者的综合作用。现在我们来研究一个有比nMOS电路。一般情况下其中上拉管电阻大约比下拉管电阻大10倍,反相器的总增益大约也是10。在此情况下,负载电容限制着上拉电压的瞬变过程,所以这个过程是比较慢的。另一方面,由高电压向低电压的瞬变过程是一种增益与负载电容的某种难以分清的混合限制作用的结果。在这种情况下,必须通过模拟才能得到精确的延迟转变时间的估算值。在这方面,CMOS电路更为理想,尽管它比单纯从阶跃响应特性预期的慢得多。图4.11为耗尽负载nMOS门的输出波形,图中各曲线分别对应于不同的输入波形。

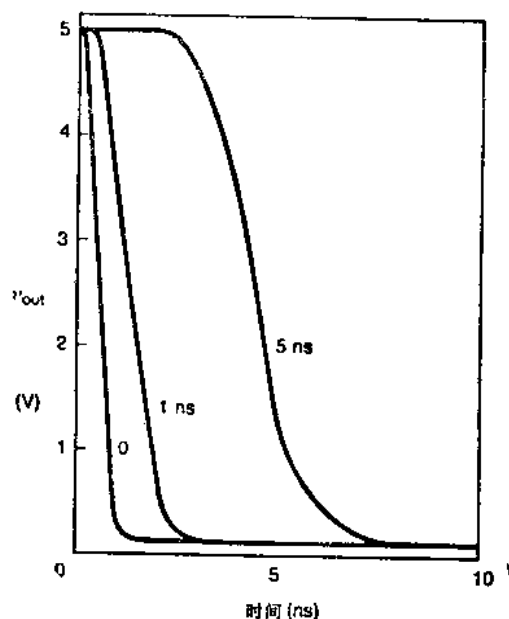


图4.11 反相器的输出波形。各斜坡输入电压20—80%的上升时间分别为0, 1和5ns

耗尽负载反相器的输入电容和前面所讨论的电阻负载反相器的相同。几种不同电路的输入充电特性示于图4.12中,它们对应于同一上升瞬变输入电压。

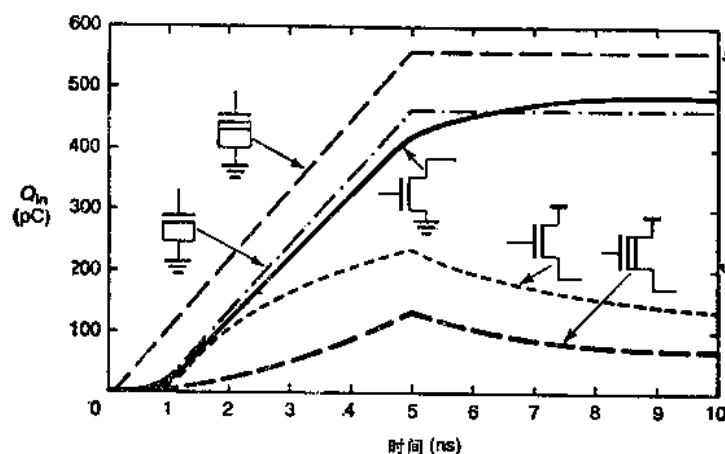


图4.12 各种电路结构的nMOS晶体管栅极总充电电荷。输入为同一斜坡电压,所有器件的尺寸相同,源跟随结构是其中最易于充电的

尽管我们对下面几级有效输入电容已作了定性的讨论,但实际上没有给出它们的定义。最合理的定义可能是

$$C_{in} = \left. \frac{dQ_{in}}{dv_{in}} \right|_{\text{对关心的区间}} \quad (4.52)$$

这种定义的问题是电荷 $Q_{in}(t)$ 可以在 $v_{in}(t)$ 停止之后还继续变化,这就使 C_{in} 失去作用。为此分析两种情况下的值:一种情况是考虑起始的 $\frac{dQ_{in}}{dv_{in}}$ 值。因为这时工作是正常的,在瞬态特性重要的起始几毫微秒期间,从驱动的角度来看,它很好地反映了负载的特性。这样一来,在超过所关心的区间时,式(4.52)还是有用的。另一个有物理意义的量是加到节点上的总电荷。从某种意义上讲,数字电路要处理的是在整个波形范围内的电荷总数,所以这个值是合理的。例如,由于密勒效应,充到增强型源跟随器的总电荷量近似等于0,源跟随器确实容易驱动。

决定了反相器的 β 比以后,下一步就是把这些结果推广到复杂逻辑门。为了作到这一点,不管实现什么样的逻辑功能,对于给定的 V_{IH} 和 V_{OL} 值,下拉网络上的压降以及流过的总电流都必须保持常数。考虑一个具有 $\frac{W_0}{L_0}$ 下拉管的典型反相器。其中下拉管的漏电压为 V_{OL} ,栅电压为 V_{IH} 。忽略短沟道和窄沟道效应,如果下拉管栅宽改为 xW_0 ,栅长改为 xL_0 ,则两个管子上将有完全一样的压降并流过完全相等的电流。这是因为两者的 S 保持不变的缘故。不论晶体管工作在线性状态、非饱和状态、饱和状态或者是截止状态,这个结论都是对的¹⁾。我们可以把晶体管分成两个串联的器件,一个栅宽为 xW_0 ,栅长为 L_0 ,另一个栅宽为 xW_0 ,栅长为 $(x-1)L_0$ 。这两个晶体管中间节点处的电压记作 V_z ,而对于沟道尺寸为 $\frac{xW_0}{xL_0}$ 的晶体管,在长度为 $(x-1)L_0$ 点上的电压同样也是 V_z 。因为两个串联晶体管中下边管子的栅宽为 xW_0 ,栅长为 $(x-1)L_0$,所以可以把它转变成栅宽为 $\frac{xW_0}{x-1}$,栅长为 L_0 的等效晶体管(以便保持 S 为常数)。

这些变换都在图4.13中说明。把这个论点一般化,则可以证明,若限定下拉管电阻的方块数保持不变,则输出电压将维持在 V_{OL} 值。注意这个论点与晶体管的电压非线性无关,而且对于任一 γ 值都成立。假设不存在短沟道或窄沟道效应,所有晶体管的栅长 L 都保持为常数 L_0 。在计算上忽略短沟道效应的作用,当 W 保持足够宽时,窄沟道效应

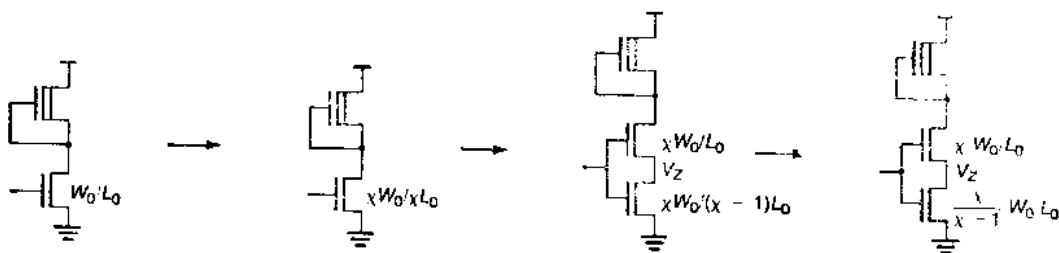


图4.13 与非门的尺寸变换。所有情况下 V_{OL} 是相同的

1) 当速度明显饱和时,这个论点就不对了。因为这时 i_{DS} 不再正比于 S 。但是对于低 V_{OL} 电压情况,这一点往往是不成问题的。

也是不重要的,因而只需要计算方块数就可以估计下拉网路。

我们知道晶体管的栅长和栅宽随加工工艺而改变,所以必须确定要用哪一个 W 和 L 值,这肯定不是版图上的栅长和栅宽,因为上面都是用有效栅长和栅宽来讨论的。可以证明在所有不同的栅长和栅宽中,从噪声角度来看有一组是最坏的。最坏情况下的噪声参量将在4.4节中讨论。

上述推导都和 V_{th} 值无关,实际上所有下拉管无例外地都与 V_{th} 有关。如果所有下拉管的输入电压都不相同,在估算所要求尺寸时就得引用线性化的方法。由于下拉管并不是完全线性的,所以线性化方法是不太严格的。然而这种近似并不太坏。比方说,对于一个给定的上拉管,当该反相器的输入信号直接来自耗尽型负载反相器时,相应下拉管所需栅宽为 W_D ,而当输入信号是通过一个传输管加进来时,相应下拉管所需栅宽为 W_P 。这就是说,我们研究的是同时存在强弱两种高电压的情况。假如有一个二输入与非门,它的两个输入端都受耗尽型负载门的强力驱动,可以预计,两个下拉管都要求 $2W_D$ 的栅宽。如果两个输入信号都是通过传输管加进来的,可以预计,两个下拉管都要求 $2W_P$ 的栅宽。我们假定一个如下的模型:令 R_D 为输入直接驱动情况下,下拉管每一方的电阻值, R_D 值的选择是任意的。按照前面讨论过的按比例变化的原则,任一下拉管应具有电阻值 $\frac{R_D L_0}{W_D}$ 。这样一来,对于输入电平不够高的二输入的与非门,如果强

驱动管的下拉电阻为 αR_D ,那么弱驱动管的下拉电阻就必须是 $\left[\left(\frac{L_0}{W_D}\right) - \alpha\right] R_D$ 。强驱动管的栅宽必为 $\frac{L_0}{\alpha}$,传输驱动管的栅宽必为 $\frac{W_P}{(1 - \alpha W_D/L_0)}$ 。

在讨论折衷方案、优化策略以及进行延时的直观计算时,建立一个MOS逻辑门延时的简单一级近似模型往往是很有用的。我们以前说过,在阶跃输入电压的情况下,通过一个MOS逻辑门对一个线性负载电容充电,所要求的时间可以写成

$$T = C \int -\frac{1}{i} dv \quad (4.53)$$

为了进行简单的延时估算,可以定义

$$R = \int -\frac{1}{i} dv \quad (4.54)$$

这里的 R 是为 RC 延时计算而定义的逻辑门等效电阻。上拉管和下拉管 R 是不同的。对于耗尽型负载反相器,可有

$$R_{pu} \approx \beta_i R_{pd} \quad (4.55)$$

R 反比于晶体管的栅宽 W 。定义 R_i 为版图上一方的下拉管电阻,这样做是很方便的。如果把管子画成 $10\mu\text{m}$ 宽,在我们的 $2\mu\text{m}$ nMOS耗尽型负载工艺中,等效动态电阻就是 $R_i/5$ 。显然,这是很粗糙的,但是却简单实用。下拉管也有等效的输入电容,这个电容正比于栅宽 W 。如果我们定义 C_i 作为具有最小栅长和最小栅宽的下拉晶体管的有效输入电容,则尺寸为 $10/2$ 的下拉管的有效电容近似为 $5C_i$ 。对于任意栅宽的下拉管, R_i 和 C_i 的乘积是常数。我们定义一个工艺常数 τ 如下:

$$\tau \equiv R_i C_i \quad (4.56)$$

为了尽可能地精确些, R_i 应当从典型电路结构的延时模拟中提取出来。 C_i 通常取为 $L^2 C_{ox}$,此处 L 是最小的版图线宽。对于n沟晶体管, R_i 的典型值是 $5-50\text{k}\Omega/\square$ 。对于

p 沟晶体管, R_s 的典型值是 $10\sim 100\text{k}\Omega/\square$ 。

CMOS反相器

传统的 CMOS 反相器示于图 1.34 中。它由一个 p 沟上拉管和一个 n 沟下拉管组成。两个晶体管都是增强型器件, 它们的栅是连接在一起的。这种电路的直流传输函数有三个工作区, 在第一区中, 两个晶体管中有一个管子是关断的。当 $v_{in} > V_{DD} + V_{TP0}$ 时上拉管关断 (对增强型 pMOS 管 V_{TP0} 是负的), 而当 $v_{in} < V_{TN0}$ 时下拉管关断。 V_{TP0} 是源端接 V_{DD} 时 p 沟管的阈值电压, V_{TN0} 是源端接 0 时 n 沟管的阈值电压¹⁾。在第二个

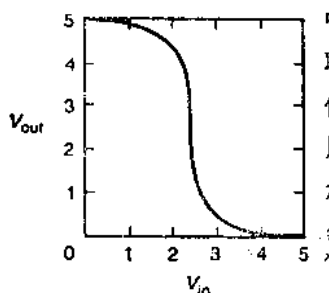


图 4.14 CMOS反相器的传输函数

重要的工作区中, 两个管子都处在饱和状态。在这个区域中, 反相器的增益非常高。增益值受限于沟道长度调制、串联电阻和漏感应势垒降低。图 4.14 给出了 CMOS 反相器的传输函数。注意由于两个管子的源端分别和相应衬底短接, 所以在这种电路结构中体效应是极其次要的。可以计算使反相器进入这个区的输入电压值。令两个管子的饱和电流相等, 可有

$$V_{inv} = V_{TN0} + \frac{V_{DD} - V_{TN0} + V_{TP0}}{1 + \sqrt{\beta \frac{K'_N}{K'_P} \frac{1 + \delta_P}{1 + \delta_N}}} \quad (4.57)$$

可以看出, V_{inv} 主要决定于管子的阈值电压, 其次才是 β 比。对于一般的 CMOS 工艺体效应是大的, 计算时必须包含有 δ 值。在 V_{inv} 点, CMOS 反相器包括有两个串联的电流源。输入电压的微小变化也会引起输出电压大幅度变化。对于一级近似, 反相器高增益区的两个端点, 就是两管之一脱离饱和状态的那两点。这两点的输出电压是

$$\frac{V_{inv} - V_{TN0}}{1 + \delta_N} \quad (4.58)$$

$$\frac{V_{inv} - V_{TP0}}{1 + \delta_P} \quad (4.59)$$

在第三个工作区中, 一个晶体管处在饱和状态另一个处在非饱和状态。

如 4.1 节所述, 起码有两个常用办法可以用来确定 CMOS 电路的电压逻辑电平。首先考察其中最简单的办法。按照在 nMOS 电路中曾采用过的方法, 把阈值电压作为参考电压, 可以定义 $V_{IL} \equiv V_{TN0}$ 和 $V_{IH} \equiv V_{DD} + V_{TP0}$ 。在这种情况下, 从噪声容限角度来看, CMOS 逻辑门中晶体管的尺寸是完全任意的。然而必须小心, 一定要有适当高的阈值电压以便得到足够的抗噪声水平。这类方法适用于中等和低等性能的芯片。

CMOS 电路设计的另一个通用方法是用逻辑门的 $-M^{-1}$ 斜率点去定义噪声容限。显然, 可以采用这个方法设计牢靠的 VLSI, 在这种设计中没有严格的理论可循。现在采用的技术是各式各样的。其基本点是保持大多数门的 V_{inv} 靠近 $V_{DD}/2$ 点, 而且要有反相器跟在预充级的后面。

与静态 nMOS 电路不同, 传统的静态 CMOS 电路只有在开关动作时才流过电流。

1) 在大多数 CMOS 工艺中, n 沟 MOS 管的衬底是接地的, p 沟 MOS 管的衬底是接 V_{DD} 的, 在 nMOS 电路中专用的衬偏发生器, 在 CMOS 电路中通常是不用的。

这个电流有两个分量。其一对外部负载电容充放电作有用之功，另一个电流分量则为上拉管和下拉管同时导通时的交叠电流。在开关频率为 f_s 时对电容 C_{LOAD} 充放电所要求的功率是 $f_s C_{LOAD} V^2$ ，其中 V 是电容上的高低电压差。我们曾经假定，加在电容上的这个能量在每个周期中都被耗散掉了。在数字电路中 V 往往等于 V_{DD} 。图 4.15 表示一个理想逻辑门。若全部电流都对电容充电，则每个周期充电电流的积分可由下式给出：

$$\int i dt = Q = CV_{DD} \quad (4.60)$$

充电电流必须从电源流出来，从电源抽出的能量为

$$\int (iV_{DD}) dt = CV_{DD}^2 \quad (4.61)$$

能量的一半耗散在 R_{pu} 上，另一半耗散在 R_{pd} 上，而和它们的数值无关。

可惜，即便是没有输出电容连接到反相器，也还是有开关能量损耗的，除非输入开关动作是瞬变的。输入变化越慢，反相器处在双管导通区的时间越长，因而超额功耗越大，最大超额电流由下式给出：

$$I_{\max} = \frac{K'_N S_N}{1 + \delta_N} \left(\frac{V_{DD} - V_{TN0} + V_{TP0}}{1 + \sqrt{\beta \frac{K'_N}{K'_P} \frac{1 + \delta_P}{1 + \delta_N}}} \right)^2 \quad (4.62)$$

CMOS 电路中的超额功率耗散是非常重要的，它很容易达到与功耗 $f_s CV^2$ 同样的数量级。这两种类型的功耗都与频率 f 成正比。 T_{out} 是输出的开关时间。开关能量 E 有两个成分。只有在放电瞬间二者才合在一起。可有

$$E = E_{\text{excess}} + E_{\text{cap}} \quad (4.63)$$

其中 E_{excess} 是相应于二管交叠电流的能量，存储在电容上的能量 E_{cap} 为：

$$E_{\text{cap}} = 1 \leq 2 C_L V_{DD}^2 \quad (4.64)$$

而

$$E_{\text{excess}} = (Q_T - V_{DD} C_L) V_{DD} \quad (4.65)$$

求解方程，可有

$$E = Q_T V_{DD} - 1 \leq 2 C_L V_{DD}^2 \quad (4.66)$$

对于快速变化的输入信号， $Q_T \approx C_L V_{DD}$ ，因此我们有

$$E \approx 1 \leq 2 C_L V_{DD}^2 = E_{\text{cap}} \quad (4.67)$$

对于变化非常慢的输入信号， C_L 是不重要的，而由反相器本身决定它的传输函数。这时 $\frac{T_{in}}{T_{out}}$ 是常数。总电荷 Q_T 正比于 $I_{\max} T_{in}$ 。在 V_{TN0} 和 $|V_{TP0}|$ 比较小的时候， $I_{\max} \propto V_{DD}^3$ ，所以在此区中

$$E \propto V_{DD}^3 T_{in} \quad (4.68)$$

(对于漂移速度饱和严重的器件， $E \propto V_{DD}^4 T_{in}$)。用电流控制的电流源驱动一个电容，并始终记录电源提供的充电电荷，就可以模拟 CMOS 电路中的动态功耗。

有一种称作“近 nMOS”的静态 CMOS 型电路要耗散直流功率。其基本观点是很简单的。因为 n 沟道 MOS 管比 p 沟道 MOS 管优越，所以我们大都愿意采用 n 沟道晶体管。

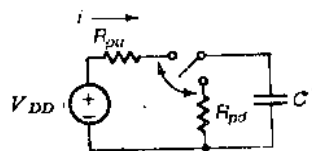


图 4.15 CMOS 反相器的理想模型。一半功率耗散在 R_{pu} 中，一半耗散在 R_{pd} 中

近年来nMOS下拉管的负载改进得越来越好。可以把p沟道晶体管看作较好的负载。图4.16(a)表明一个栅接地的pMOS上拉管，它可以很好地模仿理想的耗尽型负载（无体效应）。例如，这种类型的器件在大规模静态CMOS或非门中得到了应用。

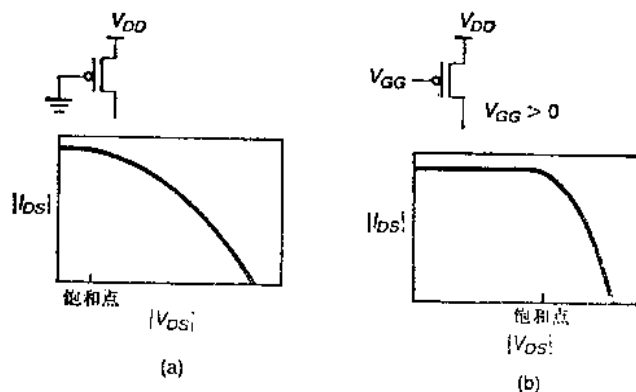


图4.16 P沟负载管。(a) $V_{GG} = 0$ ；(b) V_{GG} 接近 $V_{DD} + V_{TPO}$ 。
图(b)中的负载线比图(a)的更为理想

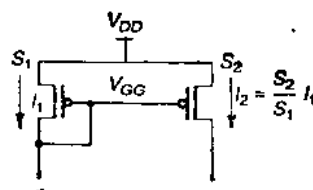


图4.17 电流镜。此电路只能在两管都处在饱和状态时工作

可以有两种办法改进栅接地的P沟道上拉管。图4.16给出两种P沟负载管。由图可见，当P沟晶体管的栅电压 V_{GG} 上升时，负载线就变得更为理想（为了保持同样电流，上拉管的栅需要更宽一些）。此外，可以通过采用称之为电流镜的常用电路技术，使流过负载的电流跟踪下拉管中的电流强度。图4.17是一个电流镜。假定晶体管 M_1 导通，它必定处在饱和状态。其电流完全决定于 $V_{DD} - V_{GG}$ 、器件形状因子和本征晶体管参数¹⁾。对于一级近似，该电流和 V_{DS} 无关。特别是在版图上器件的栅极较长时更是如此。由图看出，晶体管 M_2 的栅源电压和晶体管 M_1 的栅源电压是一样的。这样一来，当晶体管 M_2 处在饱和状态时，流过它的电流就是

$$I_2 = \frac{S_2}{S_1} I_1 \quad (4.69)$$

其中 I_1 是流过晶体管 M_1 的电流。

在MOS工艺中，不同芯片上晶体管的参数是不一样的。为了设计理想的负载，我们要求当下拉管加强时上拉管也被加强，当下拉管减弱时上拉管也被减弱。换句话说，希望下拉管电阻对上拉管电阻之比保持常数而与加工的不确定因素无关。电流镜给我们提供了一个接近达到此要求的方法。关键是要构成一个电路，由该电路产生 V_{GG} 作为参考电压，并且送到芯片各处。图4.18说明了这个电路。如果该电路中的下拉管是强的，则参考晶体管也是强的。如果参考下拉晶体管是强的， V_{GG} 就会被拉低一些，从而使芯片上各处的上拉管更难导通。

功耗是电路的一个重要参量。通常，有比门中的电流（假定输出被拉低了）正比于上拉管的迁移率。例如，对栅接地的P沟道上拉管[如图4.16(a)所示]作灵敏度分析，可有

1) 请记住符号 V_{DD} 乃是从nMOS电路设计遗留下来的一个错误符号。在CMOS电路中 V_{DD} 实际上是P沟上拉管的源端。用符号 V_{GG} 代替 V_{DD} 就更不合适（“C”是表示集电区）。 V_{SS} 通常用来表示地端。

有引起性能恶化。

采用这种技术的一个问题是高噪声灵敏度。因为 V_{GG} 送往芯片各处，所以它成为一个噪声渠道。可以预计会有大量的信号耦合到时钟线。当 V_{GG} 接近 $V_{DD} + V_{Tn0}$ 时， I_{pn} 对于 V_{GG} 噪声， V_{DD} 摆动和片内阈值电压的变化是极端敏感的。电流对于 V_T 的灵敏度为

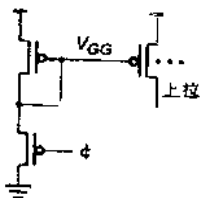


图 4.20 用于低功耗电路的电流镜和时钟参考源组合电路

$$\frac{dI}{I} = \frac{2 dV_T}{V_{GS} - V_T} \quad (4.78)$$

例如，对于 $dV_T = 100\text{mV}$ ， $V_{GS} - V_T = 1\text{V}$ ，可以看到，仅仅由于此效应，上拉管中的电流就要变化 20%。

4.3 噪 声

可用下述三个变量描述 VLSI 芯片的噪声性能：电压噪声容限、阻抗水平以及噪声源特性。在集成电路中这几个因素都对噪声有所贡献。模拟电路设计者所熟悉的某些物理噪声源，如热噪声和 $1/f$ 噪声 [Stone 84]，在数字 MOS 电路中并不重要。数字集成电路中的大多数噪声来自那些原则上可以预测的现象，但在实际中，要对它们进行系统地计算却相当复杂。

电容耦合

电容耦合是引起噪声的重要原因。当 MOS 工艺采用多层布线和连线尺寸的长宽比较大时，这种耦合变得更为重要。考虑 VLSI 中任意两个节点 A 和 B，其间耦合电容为 C_c ，如图 4.21 所示。我们希望确定节点 A 的电压 V_a 的变化对节点 B 电压 V_b 的影响，如果节点 B 和系统中所有其它节点之间总的小信号导纳（包括 C_c ）为 $Y(s)$ （其中 s 为复数频率），则 $V_a(s)$ 和 $V_b(s)$ 之间的传输函数由下式给出：

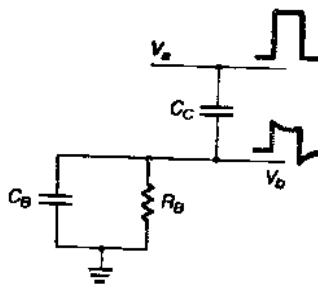


图 4.21 电容耦合噪声

$$V_b(s) = \frac{sC_c}{Y(s)} V_a(s) \quad (4.79)$$

由此可见 $Y(s)$ 的数值越大，耦合噪声越低。有两个重要的特殊情况。其一是当节点 B 用于动态存储，如果节点 B 对地电容为 C_b ，则

$$v_b = \frac{C_c}{C_c + C_b} v_a \quad (4.80)$$

第二种情况是节点 B 相应于恢复逻辑门的输出端，这时 $Y(s)$ 中含有很大的电阻分量。对于 $R_b \ll \tau / C_c$ ，可有

$$V_b(s) = sC_c R_b V_a(s) \quad (4.81)$$

因为在 nMOS 电路中上拉管的小信号电阻大约比相应下拉管的小信号电阻大 β 倍，所以

在这个效应为主的情况下，我们需要 $NM_H \approx \beta NM_L$ 。然而，有更有力的理由要求 NM_H 大于 NM_L 。在 nMOS 电路中，通常把连线预充为高电平。这时，由于上拉管关断而使它的阻抗非常之大。因此，这种情况下要求具有我们所能达到的最高噪声容限。从某种意义上讲，对此我们可以得到较大的噪声容限，这是因为高电平噪声容限是 $V_{DD} - V_{TBO}$ 的百分之几十。低电平噪声容限则是 V_{TBO} 的百分之几十，而 V_{TBO} 是一个很小的数。

电阻耦合

在 MOS 集成电路中，对噪声贡献最大的或许是来自地线和电源线的电阻反馈。由于封装，还由于晶体管的源和漏必定是 n^+ 区或 p^+ 区，集成电路的电源往往部分地通过扩散区馈送。通常扩散区具有每方几十欧姆的方块电阻，这种电阻会改变上拉管和下拉管的电阻，从而影响一个孤立的门的工作，并因此而改变了开关延时、 V_{OL} 值和功耗，这种一个门对另一个门的影响就引起了噪声。当一个门开关时，它可以改变对第二个逻辑门的有效 V_{DD} 或有效地电压。图 4.22 所示是一个电阻耦合起重要作用的例子。

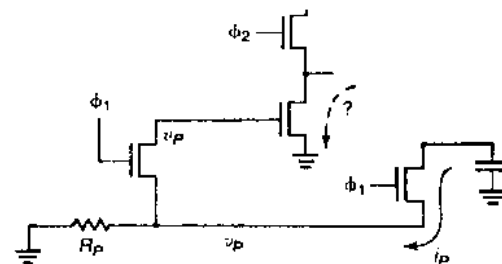


图 4.22 电阻耦合噪声。在 R_p 上产生的噪声电压，可以使在 ϕ_2 时被预充的节点偶然放电。在 CMOS 和 nMOS 电路中都会见到这种效应

在两个逻辑门具有共同的寄生电阻 R_p ，且通过它把逻辑门连到电源电压时，

如果 R_p 远远小于晶体管小信号阻抗的实数部分，则该电阻上的噪声电压 v_p 就近似等于 $i_p R_p$ 。 i_p 是由于一个门造成的流过 R_p 的电流。

在电源线中电阻导致的电压降，不仅是由扩散区电阻而且也是由接触孔的接触电阻、多晶硅线的电阻以及金属线的电阻造成的。因为金属线通常要对很多个门供电，我们可以认为：在有比 nMOS 电路设计中，除了如收缩阵列这种可能的例外外，根据统计，导线中的交流电流分量远远小于直流电流分量。系统的软失效是否是由于许多门的晶体管偶然地同时跳变而造成的，这一点目前还不太清楚。对于含有 100000 个或更多个晶体管的芯片，金属导线的电阻可能是一个重要的噪声源。

阵列结构中的噪声

由于噪声的相干性以及密集的封装密度，使得阵列结构对于耦合噪声特别敏感。当相邻连线相互跨越时引起电容耦合。图 4.23 是一个带有三种耦合电容标记的或非门阵列。减小这些有害电容耦合效应的一个方法是使电压摆幅小一些。另一个方法就是在两条耦合线之间加一个屏蔽导体。这种方法适用于并行布线和上下重叠的布线。例如，第一层金属的接地薄层可以隔离多晶硅线和第二层金属之间的耦合。

由于公共地而存在的电阻耦合或 V_{DD} 的电流也可以引起噪声。在阵列中另外一个引起噪声的重要原因是信号线或控制线间的交叠。例如，在 ROM 或 RAM 阵列中同时有二个不同的字线被选中，这种转换条件是常有的，尽管它在理想情况下是被禁止的。

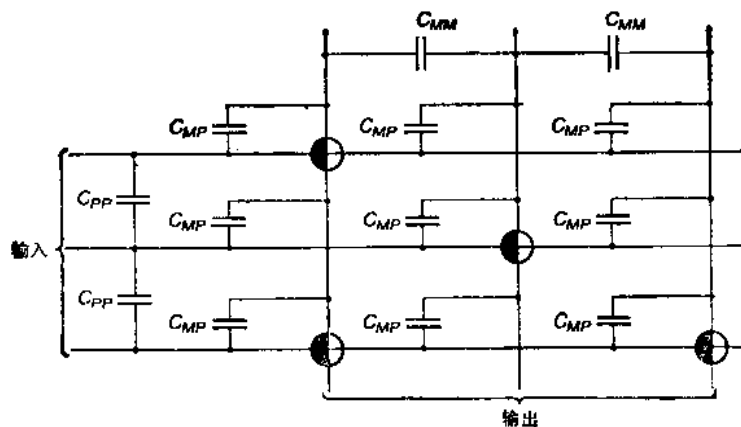


图 4.23 阵列中的电容耦合

例 4.1 HROM 中的耦合噪声。

在示例 2.9 中，指定了 ROM 的 SPICE 模型，并通过一个 nMOS HROM 分析了它的延时。现在通过了解下面的问题，把该分析扩展到含有耦合噪声的情况。假定使位线预充到 V_{DD} ，当和一条未编程的位线最邻近的两条位线的电平变化通过 5 V 时，这条未编程的位线电压要下降多少伏？

最坏的条件是当位线电容为低电平而耦合电容为高电平时，若位线互相靠得很近，则位线间的耦合电容上升比位线对地电容上升得更快，所以位线宽时为最坏条件。因为位线是按照 $4\mu\text{m}$ 线宽和 $4\mu\text{m}$ 间距来画的，如果金属线条误差在最坏条件下为 $\pm 0.4\mu\text{m}$ /每边，则最终线宽为 $4.8\mu\text{m}$ ，间距为 $3.2\mu\text{m}$ 。

我们先求出每微米位线的平均对地电容。若重复间距为 $16\mu\text{m}$ ，对于未编程位线，这包括了 $C_{\text{midiff}}144\mu\text{m}^2$ $C_{\text{mpot}}19.2\mu\text{m}^2$ ， $C_{\text{mf}}43.2\mu\text{m}^2$ 。由此得到平均位线电容为 25.6×10^{-1} pF/ μm 。从 Dang 的文章 [Dang 81] 可知，若金属线平均高出硅表面 $1.1\mu\text{m}$ ，其厚度为 $0.55\mu\text{m}$ 。则总的耦合电容为 2.5×10^{-5} pF/ μm 。因而耦合只下降约 21dB，这就是说，在邻近位线上的电压变化 5 V，将会导致在未编程位线上有 460mV 的电压变化。在最坏条件下，即便是该位线已被充分编程，电压仍将变动大约 300mV。

α 粒子噪声

在含有 4000 位或更高位存贮电容的动态存贮器中，曾经发现 α 粒子是一个重要的噪声源。当微处理器处于增强辐射的条件下，也曾观察到与 α 粒子相关的软失效。

α 粒子是一个双重离化的 He 原子，它所具有的能量约小于 10MeV。这些 α 粒子穿透硅的厚度高达 20 或 $30\mu\text{m}$ ，并在它穿过的路径上产生电子-空穴对，每微米路径上的电子-空穴对有 25000 到 80000 个。当这些电子和空穴被收集时，它们可形成噪声电流，或者更重要的是形成噪声电荷。显然，设计一个具有低收集效率的动态 RAM 单元是非常重要的。

尽管在封装方面进行了改进，但是要使 α 粒子通量低于 0.001 到 $0.01\alpha/(\text{cm}^2 \cdot \text{h})$ 似乎是不大可能的。因而良好的电路设计和系统设计是很重要的。通常对存贮单元确定一个临界电荷 Q_{crit} ，作为存贮单元对 α 粒子灵敏度的一级表示，然而，还有很多其它因素，

像收集面积等, 不管设计细节如何, N_{crit} 超过约 2×10^6 个电子的存贮单元看来是十分安全的。低于这个水平, 就要求细致设计。对于良好的设计单元, Q_{crit} 的大小可以降低一个数量级。假定有一个 $W \times W$ 的方块单元, 可有

$$Q = \Delta V_{crit} C_{ox} W^2 \quad (4.82)$$

或

$$W_{crit} = \sqrt{q N_{crit} / \Delta V_{crit} C_{ox}} \quad (4.83)$$

对于 $\Delta V_{crit} = 1 \text{ V}$, $C_{ox} = 10^{-3} \text{ pF}/\mu\text{m}^2$ 和 $N_{crit} = 2 \times 10^6$, 可以得到 $W_{crit} \approx 18 \mu\text{m}$ 。虽然我们假定的 N_{crit} 值是保守的, 但是很清楚, 在动态逻辑电路中 α 粒子的噪声不是不可能存在的。分析上述例子可以得到最小的 $C_{crit} = 0.3 \text{ pF}$, 按 VLSI 电路标准, 这是一个十分大的电容 (通过细心地设计, 实际上 DRAM 中的 C_{crit} 值接近于 0.05 pF)。我们再从另外一个角度考察一下这个问题。一个 $3 \mu\text{m}$ 宽的耗尽层, 在 α 粒子最严重的情况下, 可以有 200000 个电子-空穴对, 所造成的电流脉冲的峰值约为 1 mA [May 79]。当 VLSI 逻辑芯片中的动态存贮节点数上升时, 与 α 粒子相关的软失效显然就会成为问题。

工艺加工偏差引起的噪声

参量不一致也会形成噪声分量。尽管芯片之间及硅片之间器件参量是不同的, 但一般还是假定任意一个芯片上的器件参量是完全一致的。在这一假定不再成立的时候 (当芯片上器件数目增加时, 该假定的真实程度就更加降低), 这些参量的不一致性就引起噪声。例如, 我们考虑 V_{TE0} 的变化。该变化可看成是与下拉晶体管串联的一个电压噪声源。这就是说, 除去该噪声的频谱包括一个直流冲击以外, 它和其它任何感应噪声电压是一样的。

为了弄清随着工艺的变化芯片上的物理参数 (如多晶硅线宽) 是怎样转化为噪声电压的, 我们看一个简单的例子。在有比耗尽型 nMOS 电路中, 输出低电压决定于上拉管的形状因子对下拉管形状因子之比。下面我们只考虑输入高电压接近 V_{DD} 时的一级效应。假定 $K'_n \approx K'_p$, 且下拉管处在线性区, 则有

$$V_{out} \approx \frac{V_{TD0}^2}{2\beta(V_{in} - V_{TE0})} \quad (4.84)$$

我们集中分析下拉管栅长 L 的变化并考查 dV_{out}/dL 值。可以得到

$$\frac{dV_{out}}{dL} = -\frac{\partial V_{out}}{\partial \beta} \frac{\partial \beta}{\partial L} + \frac{\partial V_{out}}{\partial V_{TE0}} \frac{\partial V_{TE0}}{\partial L} \quad (4.85)$$

或

$$\frac{dV_{out}}{dL} = V_{out} \left(\frac{1}{L} + \frac{1}{V_{in} - V_{TE0}} \frac{\partial V_{TE0}}{\partial L} \right) \quad (4.86)$$

其中 $\frac{\partial V_{TE0}}{\partial L}$ 可从短沟道模型得到, 或者为了更加保险可从实验数据中取得。我们希望把 L 的变化和 V_{in} 的变化联系起来。可有

$$\frac{dV_{out}}{dV_{in}} = -\frac{V_{out}}{V_{in} - V_{TE0}} \quad (4.87)$$

求解 dV_{in} 得到

$$dV_{in} = - \left(\frac{V_{in} - V_{TBO}}{L} + \frac{\partial V_{TBO}}{\partial L} \right) dL \quad (4.88)$$

从式(4.88)可以看出, 对于 $V_{in} \approx 5\text{ V}$, $V_{TBO} \approx 1\text{ V}$, 在不包括短沟道效应的作用时, L 变化10%可引起400mV的等效直流噪声电压。对于质量控制这是一个很好的依据。

例4.2 工艺偏差引起的噪声。

假定增强型晶体管阈值电压的工艺分散性是芯片上唯一的噪声源, 而且可用一个正态分布恰当地描述它的统计起伏。对于一已知具有 N 个下拉管的电路, 求解噪声电压在某处超过噪声耐度的几率是多少? 如果所有的噪声电压都小于噪声容限的几率为 p , 若假定了所有噪声电压是相互独立的, 则任意一个噪声电压低于此极限的几率必为 $p^{1/N}$ 。

当 $p \approx 1$ 时, 我们可以把它近似为

$$p^{1/N} \approx 1 - \frac{(1-p)}{N}$$

在正态几率分布 $N(X)$ 的尾部区可以近似为

$$[1 - N(X)] \approx \int_X^\infty \frac{1}{\sqrt{2\pi}} e^{-(1/2)y^2} dy \approx \frac{1}{\sqrt{2\pi} X} e^{-(1/2)X^2}$$

其中 X 是以 σ 归一的, σ 为标准偏差, 并有

$$\frac{(1-p)}{N} \approx \frac{1}{\sqrt{2\pi} X} e^{-(1/2)X^2}$$

或

$$X \approx \left\{ -2 \ln \left[\sqrt{2\pi} X \frac{(1-p)}{N} \right] \right\}^{1/2}$$

对于 $p = 0.99$ 和 $N = 10^5$, 在这个示例中有

$$X \approx 5.2$$

或

$$\sigma_{V_{TBO}} \leq \frac{NR_L}{5.2}$$

NR_L 是低电平噪声耐度。与之比较, 晶体管的95.6%阈值电压变化要在标定值的 $\pm 2\sigma_{V_{TBO}}$ 之内。

4.4 最坏条件设计

我们已经讨论了集成电路中噪声的起因和影响。在某些情况下, 噪声在本质上确实随机的。而在另一些情况下, 噪声是由某些现象所产生的。尽管这些现象是可以预计的, 然而, 对它们进行精确计算却太复杂了。例如, 通过寄生电阻和寄生电容的信号耦合就属于这一类。

工艺的偏差造成了许多附加的尺寸不确定性。在批与批之间、硅片与硅片之间、芯片与芯片之间乃至器件与器件之间, MOS晶体管的栅长、栅宽、杂质分布、电子迁移率等参数, 都有很大变化。幸好, 这些参数的控制, 对同一芯片上的晶体管比对不同芯片上的晶体管, 要好上一个数量级。把工艺参数的偏差分为两个部分是有益处的。一部分偏

差在芯片范围内是不变的,但不一定在整个硅片范围内是不变的。第二部分偏差是使所有参数的变化都集中于单个芯片上,后者可以简单的看做是另一个噪声源(参看4.3节)。

需要说明的是设计集成电路要求在整个给定的温度范围内满足所有的功能规范以及保证所有的噪声容限。我们所处理的器件参数变化范围是很宽的,然而,在一个芯片上它们是常数。这是因为我们已经分解出了非常数部分,以使它们通过噪声容限参加运算。参数 T_{ox} , V_T , ΔW 和 ΔL 在整个芯片上应假定为常数。恰当地考虑器件参数间的相互关系对设计和制造高性能电路是很重要的。

选择何种合适的电路设计方式是随着市场对芯片的需求而定的。一般讲,系统市场要求部件能满足某些最坏条件下的规范,而在元件市场中,往往把元件入库或分级,速度较快的元件可以卖更多的钱,但大部分低速元件也不丢掉。在某些条件下,异步系统也可能属于上述第二类。

显然,评价加工所得芯片的方法对设计方式有影响,因为这种设计方式是实现最高评价所需的。我们将考虑适用于大多数系统的共同情况,即芯片运行速度非常快但过热,这种情况与芯片温度很低但速度太慢同样是没有实用意义的。在这些限制之下,对统计上有特殊意义的“最坏”条件进行设计是很重要的¹⁾。这就是说,芯片应该设计成在最坏条件下的低速芯片,速度足够快,而最坏条件下的高速芯片,温度足够低,同时在一切实情况下逻辑必须具有正的噪声容限。

为了对这个问题得到一些直观的知识,我们考查一个包含有1000个反相器的普通LSI。假设每个反相器可在100ns内对1pF电容充电,它所消耗的直流功耗为1mW。即功耗延时乘积为100pJ。进一步假定迁移率是唯一重要的工艺参数,其额定值为1000,变化范围是670—1500。这样一来,在工艺偏差允许的一个极端情形下,芯片消耗功率1.5W,并可在67ns内对电容充电。在另一个极端情况下,芯片消耗功率0.67W,同时在150ns内对电容充电。在这两种情况下,功耗延时乘积都是100pJ。然而要注意,在必须满足最坏条件规范时会发生些什么。假定芯片功耗必须小于1W,这就意味着必须降低晶体管的形状因子,使得在最坏条件下的功耗是1W而不是1.5W。这时最佳条件下的功耗就是0.44W,最佳速度是100ns,而最坏条件下的速度则为225ns。因为我们必须对功耗最坏条件和速度最坏条件同时进行设计,所以功耗最坏条件下的功耗与速度的乘积(225pJ)对最坏条件设计是重要的参数。即使是这两个条件未发生在同一芯片上,这一点也是对的。由于不知道何种情况会发生,所以迫使人们对两种情况进行设计,好像这两种情况会同时发生似的。

因为工艺参数表现为从不合理到极合理之间是连续变化的,这就必须确定分布曲线上器件成为不可接受的极限点,这正是前面所说的什么是“统计上有重大意义”的含意。设计所能允许的参数范围越宽,成品率越高,但最坏条件的性能就越差。典型的设计做法是允许器件参数变化到 2σ 或 3σ 点处。由于在 2σ 点处可认为对于任一参数全部器件的95%是合格的,所以 N 个独立参量使成品率降低的最大倍率是 0.95^N 。这是一个较低的界限,因为即使两个独立参量超出规范,仍可以使最终电路保持在规范之内。这里的 N 是一个小数,它与芯片上晶体管的大小或数量无关,它或许是工艺复杂程度或成熟

1) 另一种设计方法是中心值设计。这种中心值设计是基于工艺参量统计分布的形状,使与工艺有关部分的成品率最高。这种更为复杂的方法适用于大批量生产,并且超出了本书的内容。

程度的一个较好度量。我们将看到，有三个原因使如何说明最坏条件参数范围的标准的确立变得非常复杂，首先统计法是复杂的；其次，对于电路设计来讲，某些参数比另外一些参数更为苛刻；第三，参数是相关的。例如， V_{T0} 值依赖于晶体管栅长和离子注入剂量的变化。用条形图表示的阈值电压典型分布示于图4.24中，这些变化是对一种最小栅长器件的统计结果。图4.25是散点图，它表示阈值电压与器件栅长之间的联系很少。

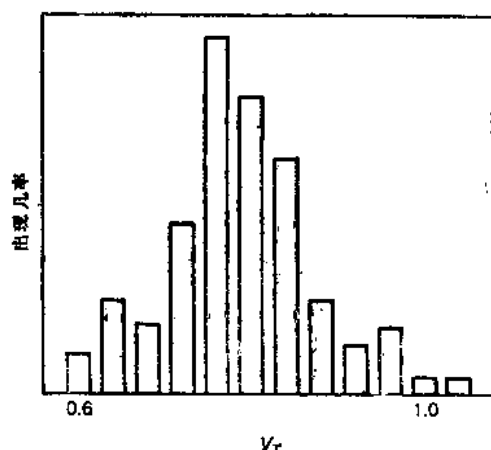


图4.24 用条形图表示的阈值电压出现的几率。有些硅片的阈值电压将会超出所允许的极限

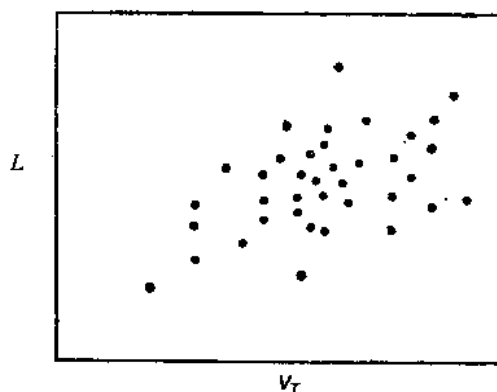


图4.25 阈值电压对沟道长度的分散图。它表示了二者之间的微弱联系

对于一个任意网络，确定什么样的参量组合产生最坏条件是非常困难的（容差问题）。由于设计方法对电路形式上的一些限制，使得确定MOS数字电路的最坏条件并不如此复杂，最坏条件设计乃是在工艺师、器件物理学家和电路设计师们的共同参与下完成的。工艺师可以把器件参数的范围和相互关系提交给电路设计师。原则上讲，工艺师不能把最坏参数交给电路设计师，这是因为在能够解释什么是“最坏”的含意之前，就必须先选定电路形式。实际上，工艺师往往把最坏条件参数表交给电路设计师，这是由于电路形式已经达到了高度的成熟和稳定。如果电路设计师要研制一个完全新的电路形式，工艺师和电路设计师就需共同评价在这个新项目中“最坏”二字的含义。最坏条件参数表常常采取电路模拟文件的格式。例如，可能有一个文件是用于噪声最坏条件，而另一个文件是用于功耗最坏条件。这些表格包含了工艺间的联系，而且暗暗假定了一组电路的形式。据称每个最坏条件文件代表了许可的工艺变化空间中的一个区域。

可以把最坏条件参数表看作是工艺师和电路设计师之间的约定。工艺师同意以合理的成品率提供硅片，其性能处于最坏条件规范的一个区域。实际存在的变化可能大于最坏条件规范给出的这些范围，但双方同意，超出规定范围的芯片就是次品。另一方面，设计师同意不利用许多潜在的关系，或者虽然利用了这些关系，至少所冒的风险由他（或她）自己负责。从式（4.29）可以知道，在阈值电压和栅氧化层厚度之间存在着一定关系。在事先没有同工艺师说清楚的情况下，设计师企图利用这个关系将是不明智的。可能有一些其它更加重要的效应。例如，由于离子注入是通过栅氧化层进行的，所以注入剂量和分布也和 T_{ox} 有关。即使没有其它的条件，工艺师可能，并多半会考虑用

这些关系,以便在合理成品率的情况下使规范得到满足。另一个重要之点是保守的电路设计师仅仅对于那些工艺师可在工艺线上检测的参量抱以期望,可以期望这些参数规范的一致性。

nMOS

对于示例中曾引用过的耗尽型负载 nMOS 逻辑系列,我们选取了 $V_{TL} < V_{TE0}$, 这就意味着 $V_{OH} = V_{DD}$ 。在输入电压为 V_{IH} 时,对于全部有效的工艺参数区间、温度和电压范围, V_{OL} 必须小于或等于 $V_{TE0} - NM_L$ 。 V_{OL} 和很多因素有关。简单地讲,对于低的下拉管与上拉管电导比, V_{OL} 值最大。在阈值电压与有效 β 比同时低的条件下,噪声容限最小(考虑最坏条件之一)。

在一种参数的不同组合情况下会出现功耗最坏条件。这就是上拉管的电导最大以及下拉管的电导为最大,而后者影响较轻。由于 nMOS 电路中功耗的主要部分来自静态,所以只有在非常特殊的情况下,才需要考虑电容充电和放电时的功耗。通常在功耗最坏条件下电路工作速度非常快。

在几乎与上述情况相反的条件下出现速度最坏条件。由于门延时的主要部分来自上拉管的上拉周期,所以上拉管电导低时出现速度最坏条件。速度最坏条件比功耗最坏条件更加复杂,这是因为门延时不仅仅依赖于上拉管而且还与负载电容有关。尽管金属和多晶硅电容基本上与有效晶体管电导无关,而栅电容和扩散区电容却与晶体管参数有关。温度也有一种一级效应,电子迁移率随温度升高而降低,所以在高温下出现速度最坏条件。应当注意,次开启漏电在高温时也要上升¹⁾。

现在我们一个参量一个参量地考虑 nMOS 数字电路中导致电流最坏条件和功耗最坏条件的条件。在功耗最坏条件下,耗尽型器件和增强型器件都是“快速工作”的。这就是说,上拉管和下拉管都是高度导通的。功耗 P 的主要部分是耗散在耗尽型负载上的。对于耗尽型负载上拉管,它的栅和源短接, V_{DS} 近似为 V_{DD} , 可有

$$P \approx K' S_{pu} V_{TD}^2 V_{DS} \approx K' S_{pu} V_{TD}^2 V_{DD} \quad (4.89)$$

在功耗最坏条件的条件下, V_{DD} 、 K' 和 W 取最大值, L 和 V_{TD0} 取最小值。由于 V_{TD0} 是负的所以它是最小值,因而从最小的 V_{TD0} 得到最大的 V_{TD}^2 。当 μ 为最大值和 T_{OX} 为最小值时, K' 也是最大值。这些都是关于功耗的一级效应。由于在温度超过约 77K 时迁移率随温度的升高而降低,所以在低温时出现电流最坏条件。尽管电源电压 V_{DD} ²⁾ 和温度在最坏条件的计算中是重要的,但它们并不是最坏条件器件模型的一部分。我们之所以关注温度,不仅仅是因为温度对迁移率的一级效应,而且还由于温度对阈值电压的二级效应。当温度降低时,阈值电压升高,实际上这将降低耗尽型器件和增强型器件的电导。这个效应与迁移率的贡献相反,其作用比迁移率的贡献要弱一些。这是一个起缓和作用的例子。功耗最坏条件的模式通常被称作“FF”或“F”模式,它代表“快速耗尽和快速增强”。在芯片温度高的时候,我们关心的是功耗最坏条件,所以功耗最坏条件发生在高温,这就是说,在芯片温度低的时候我们并不担心芯片会烧毁。除了温度以外,功耗最坏条件的参数和电流最坏条件的参数是一样的。在芯片温度最高时,出现功

1) 低温时在耗尽型晶体管中会发生杂质凝聚,所以对这种温度也必须小心从事。

2) 在 5V 工艺中, V_{DD} 的典型规范是 4.5—5.5V, 或者 4.75—5.25V。

耗最坏条件而在芯片温度最低时则出现电流最坏条件。因为迁移率随温度升高而降低，所以在功耗最坏条件下，流过芯片的电流要比在电流最坏条件下流过的电流小些。表 4.1 列出了功耗最坏条件的条件。

功耗最坏条件只是设计空间中的三个重要区域之一。速度最坏条件发生在耗尽型晶体管 and 增强型晶体管都“慢速工作”的时候（称之为“SS”或“S”模式）。虽然逻辑电平 V_{IL} 和 V_{IH} 的确依赖于 V_{TB} 和 V_{TD} 值，但是阈值电压向更正的方向变化的主要影响是降低上拉管和下拉管的电导。低温下迁移率的变化导致电流最坏条件，属于同一个原因，在高温下出现速度最坏条件。

表 4.1 导致功耗最坏情况和电流最坏情况的条件

参 量	最坏情况下的数值	原 因
τ	高（低）	
V_{DD}	高	
V_{TD}	更负	V_{TD}^2 升高
μ	高	K' 大
T_{ox}	低	K' 大
ΔW	最小	器件栅宽
ΔL	最大	器件栅短
L_{diff}	最大	器件栅短
ΔL_{poly}	最大	器件栅短
γ	低	V_T 低
V_{SB}	负的少些	V_T 低
V_{TBO}	低	

表 4.2 导致速度最坏情况的条件

参 量	最坏情况下的数值	原 因
τ	高	低迁移率
V_{DD}	低	较小驱动
V_{TD}	负的少些	较小驱动
μ	低	低 K'
T_{ox}	厚	低 K'
ΔW	最大	窄栅管
ΔL	最小	长栅管
L_{diff}	最小	长栅管
ΔL_{poly}	最小	长栅管
γ	高	高 V_T
V_{SB}	负的少些	更大电容
V_{TBO}	高	

在出现速度最坏条件时，动态工作问题也很重要。在这种情况下所有连线电容均取其最大值，而栅电容却取其最小值。这是因为延时正比于电容而反比于驱动电流。当栅氧化层厚度增加、晶体管栅宽减小时，上拉管或下拉管的有效驱动能力下降，其下降速率几乎和栅电容的下降速率相同¹⁾。然而，因为负载电容的一部分是由栅电容组成的，一部分是由连线电容组成的，所以它的下降受 $K'S$ 支配。在速度最坏的条件，栅电容和交叠电容是低的，这些都起着缓和作用。与此相同的论证适用于栅面积。对速度最坏条件， ΔW 最大，而 ΔL 最小。假设要建立一个反相器链，就会发现延时将随 V_{DD} 而上升。当 V_{DD} 升高时，电压摆幅就要高一些，这就需要更长的时间使上拉管饱和。另一方面，从整个芯片的角度来看，由于预充线和所有引线及芯片的最大时钟频率将随 V_{DD} 的降低而下降，所以最坏条件发生在 V_{DD} 低的时候。表 4.2 总结了导致速度最坏条件的条件。

在快速上拉慢速下拉和 V_{TBO} 低同时出现时（称为“FS”或“噪声”模式）产生噪声容限最坏条件。 V_{TBO} 值低和下拉弱是矛盾的。从式 (4.45) 可知，低电平噪声容限由下式给出：

$$NM_L = 2V_{TBO} - V_{IH} + \sqrt{(V_{IH} - V_{TBO})^2 - \frac{K'_D}{K'_S} \frac{V_{TD}^2}{\beta}} \quad (4.90)$$

1) 由于垂直场引起迁移率降低，这两个速率只是“几乎”相同而非完全一致。

在输入端由增强型上拉管驱动的情况下（例如零阈值的大缓冲器中的传输管），式(4.90)中的 V_{IH} 相应地为 $V_{DD}-V_{TS}-NM_H$ ，或者是 $V_{DD}-V_{TS}-NM_H$ 。如果栅极由强高电平信号驱动，我们就有 $V_{IH}=V_{DD}-NM_H$ 。低电平噪声容限与 V_{TE0} 的依赖关系是十分复杂的，从图 4.26 中可以看出这一点。

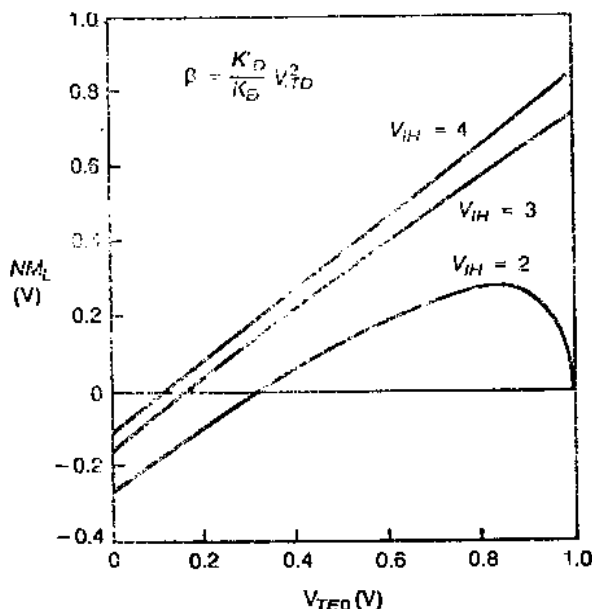


图 4.26 各种输入高电平下的低电平噪声容限图。注意，在 V_{IH} 逻辑电平下曲线不是单调变化的

考察一下 NM_L 对 V_{TE0} 的变化灵敏度，则得到

$$\frac{\partial NM_L}{\partial V_{TE0}} = 2 - \frac{V_{IH} - V_{TE0}}{\sqrt{(V_{IH} - V_{TE0})^2 - \frac{K'_D}{K'_E} \frac{V_{DD}^2}{\beta}}} \quad (4.91)$$

当

$$V_{IH} - V_{TE0} = \sqrt{\frac{4K'_D V_{DD}^2}{3K'_E \beta}} \quad (4.92)$$

时，灵敏度改变符号。

由于最小 β 比由式 (4.45) 给出，可以得

$$V_{OH} = V_{TE0} + NM_H + \frac{4(V_{TE0} - NM_L)^2}{8(V_{TS} - NM_L) - 3} \quad (4.93)$$

当电压处在这一点时，噪声容限对 V_{TE0} 的灵敏度等于 0。

式 (4.93) 暗示，对于一般的中等强度的输入高电压，低电平噪声容限对 V_{TE0} 的灵敏度与低 V_{TE0} 导致的最坏条件是一样的。在非常低 V_{IH} 值的罕见情况下（例如，从 TTL 电平转换到 MOS 电平时）， V_{TE0} 最坏条件实际上发生在 V_{TE0} 高的时候。对于非常低的 V_{IH} ，下拉管的驱动能力是很低的，因而它成为决定最坏条件的主要因素。在最坏条件下，增强型下拉管的迁移率是低的，而耗尽型上拉管的迁移率则是高的。通常在噪声容限最坏条件的文件中反映这两个量之间的联系。

为了决定在最坏条件下 ΔL 和 ΔW 的数值,我们还要进一步分析设计方法。当其它情况相同时,处理的时间常数 τ 正比于 L_{eq}^2 。这就使下拉管的栅长减至最少成为非常重要的问题。耗尽型上拉管的栅长对性能的影响不怎么严重,其原因是:它大部分时间处在饱和状态,决定逻辑门速度的电容是沟道对衬底的电容,这个电容通常比增强型下拉管栅对沟道的电容低一个数量级(在耗尽型上拉管中,栅通过源区与沟道短接)。

我们知道,有比逻辑门的速度和功耗都主要决定于耗尽型负载。当保持最坏条件的功耗为常数而改变上拉管栅宽的版图尺寸时,我们考查一下最坏条件的延时会发生什么变化。对于额定的栅长 L 和栅宽 W ,假定栅长变化是 $\pm \delta L$,栅宽变化是 $\pm \delta W$,若忽略短沟道效应,为了保持最坏功耗条件为 P_w ,上拉管的实际形状因子必须保持为常数。当实际栅宽为 $W + \delta W$,实际栅长为 $L - \delta L$ 时,可有

$$P_w = P_o \frac{W + \delta W}{L - \delta L} \quad (4.94)$$

式中 P_o 为比例常数。求解 L ,我们得到

$$L = \delta L + \frac{P_o}{P_w} (W + \delta W) \quad (4.95)$$

图4.27表示出了 W - L 平面上的等功耗线。该图也指出了设计规则的限制。图4.28表明了最坏条件下的功耗必须小于某值,表明了在这种限制条件下可能有解的区域。

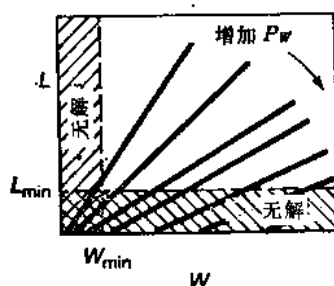


图4.27 等功耗线图

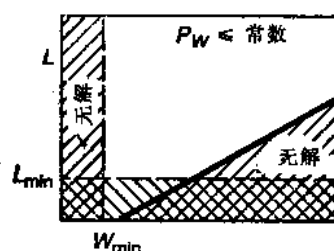


图4.28 对最大功耗限制的可能有解区

速度最坏条件有点复杂,这是因为存在着上拉管固有电容的缘故。电容公式中有四个重要项,其中有一项与 W 和 L 无关,它包括负载和隐埋接触。还有一项的大小正比于器件的实际栅宽。第三项的大小正比于器件栅长。第四项与沟道和衬底有关,它的大小正比于实际的器件面积。上拉管流过的电流正比于实际的形状因子,当实际栅宽为 $W - \delta W$,栅长为 $L + \delta L$ 时电流为最小值。综合所有这些效应可以得出

$$T_w = T_o \frac{L + \delta L}{W - \delta W} [1 + a_1(W - \delta W) + a_2(L + \delta L) + a_3(L + \delta L)(W - \delta W)] \quad (4.96)$$

其中 a_i 表示固有电容公式中的常数。 T_o 是一个比例常数。现在暂时忽略 a_1 到 a_3 项,用式(4.95)的 L 值代入,可以得知在恒定的功耗最坏条件下,尽可能地加大 W 和 L 是有利的,这是因为大的 W 和 L 值可使控制电流最大。如果把 a_1 到 a_3 相应的固有电容项包括在内时,最坏条件的延时 T_w 会出现一个极小值。对于非常小的 W 值,由于加工工艺中的不确定性,我们必须把 L 作的比希望值大一些(因而延时也大一些)。对于非常大的 W 值,加工不确定性减至最小,但是固有电容则反过来开始影响延时。图4.29给出最坏条件的延时 T_w 对 W 的典型曲线。图中并标明选择 W 的良好方案。

在设计 VLSI 芯片时, 往往不可能花时间使每个门优化到上述的细致水平。电路设计要解决的实际问题是包含成千上万个门的整体问题。由于功耗是涉及整体的, 并不是像上边讨论的受某个局部限制, 所以我们还没有分析到实际的问题。尽管如此, 式(4.96)表示的一般性能也是整体电路的一般性能。不要把上拉管的栅长和栅宽作的太小, 以便使器件工作在远离曲线 W^{-1} 的陡变部分。根据一般规则, 在耗尽型负载 nMOS 电路技术中, 我们常常把上拉管栅长作的比下拉管的长一些, 而下拉管则保持栅长为最小。例如, 在我们的 $2\mu\text{m}$ nMOS 工艺中, 从未(也可能极少)把耗尽型上拉管设计成小于 $6\mu\text{m}/6\mu\text{m}$ 。

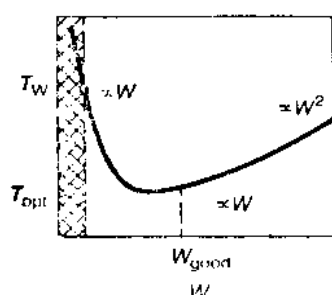


图4.29 恒定功耗时的延时与上拉管栅宽关系曲线。在低 W 值时, 线宽控制成为问题。在高 W 值时, 固有电容成了问题

表4.3 导致噪声容限最坏情况的条件

参 量	最坏情况下的数值	原 因
τ	高	$V_{T_{EC}}$ 低
V_{DD}	低	V_{OL} 高
V_{TD}	更负	V_{TD} 升高
μ_D	高	/
μ_E	低	/
T_{ox}	/	/
ΔW	最小	宽栅上拉管
ΔL	最小	长栅下拉管
L_{diff}	最小	长栅下拉管
ΔL_{poly}	最小	长栅下拉管
γ	低	低 V_T
V_{BB}	负的少些	低 V_T
V_{TE}	低(通常)	$V_{TE} - V_{OL}$ 低

现在可以回到在噪声容限最坏条件设计时如何确定最坏条件的 ΔL 和 ΔW 值这个问题上来了。由于上拉管栅长大于下拉管栅长, 噪声容限最坏条件发生在快速上拉和慢速下拉的时候, 这时 ΔL 为最小值 (L 为最大值)。在用百分比表示时, 它对下拉管作用较大。通常也限定下拉管的栅要宽一些, 或者说比上拉管的栅宽一些。因为 $\beta \gg 1$, 所以这一点是容易作到的。把这个限制加到设计方法中去, 对于噪声容限最坏条件设计, 我们就有 ΔW 最小 (W 最大)。对耗尽型负载 nMOS, 噪声容限的最坏情况的条件列在表 4.3 中。

例4.3 根据噪声容限确定反相器的大小。

选择下拉管栅宽, 使之满足噪声容限的要求, 上拉管的版图尺寸限制为 10/10 和 6/6。

规范是: $NM_H = 800\text{mV}$, $NM_L = 200\text{mV}$, $4.5\text{V} \leq V_{DD} \leq 5.5\text{V}$, $-3\text{V} \leq V_{BB} \leq -2\text{V}$ 和 $0^\circ\text{C} \leq \tau \leq 100^\circ\text{C}$ 。

在噪声最坏条件的局部区, $V_{DD} = 4.5\text{V}$, $V_{BB} = -2\text{V}$, 和 $\tau = 100^\circ\text{C}$, 最坏条件的文件是 “N” 或 “FS”。可以运行 N 模式的 SPICE 程序, 并打印出阈值电压信息。需要对 NFS 参数作出注解, 因为这是 SPICE 程序启动次开启模型的一个标志, 并使输出器件的阈值电压不确切了。在 $\tau = 27^\circ\text{C}$ 时, 对于 $20\mu\text{m}$ 栅长的器件, 得到 V_{Tx} 值为 0.576V 。温度升高到 100°C 时, V_{Tx} 下降到 0.480V 。对于栅长为 $2.25\mu\text{m}$ 的器件, 它有 8mV 的短沟道效应, 使 V_{Tx} 降到 0.472V 。加上 $V_{BB} = -2\text{V}$ 后, V_{Tx} 上升到 0.503V , 漏后

应势垒下降使 V_{TS} 降低109mV, 因而最终得到的阈值电压是0.394V。从式 (4.45)可有

$$W_{pd} = \frac{L_{pd} S_{pd} \mu_n V_{TD0}^2}{\mu_n [2(V_{TH} - V_{TS0})V_{OL} - V_{OL}^2]}$$

其中 $V_{TH} = 4.5 - 0.8 = 3.7V$, $V_{OL} = 0.394 - 0.200 = 0.194V$ 。在室温条件下, $\mu_p/\mu_n = 730/670 = 1.09$; 在100℃时, 这个比值不变。 $L_{pd} = 2 + 0.25 - 0.38 = 1.87\mu m$ (0.38 这一项是由于横向扩散引起的)。对于版图尺寸为10/10的上拉管, 形状因子是8.8/9.87; 对版图尺寸为6/6的上拉管, 形状因子是4.8/5.87, 因此在10/10的情况下, 可有

$$W = \frac{1.87 \times 0.89 \times 1.09 \times (-2.3)^2}{[2 \times (3.7 - 0.5) \times 0.19 - 0.19^2]} = 8.13$$

或者

$$W_{drawn} = 9.3$$

对于6/6的上拉管, 可以得到 $W_{drawn} = 8.6$ 。要注意 V_{TS} 有两个不同值: $V_{DS} = 0V$ 时 V_{TS} 为0.503V; $V_{DS} = 5V$ 时, V_{TS} 为0.394V。 $V_{DS} = 5V$ 时的 V_{TS} 值用来计算所需的 V_{OL} 值。

因为有许多二级效应(如垂直场引起某些参数退化等)都没有考虑进去, 所以以上这些数值只能是初步的猜测。从SPICE 程序我们得到: 对于10/10的上拉管, 下拉管的版图尺寸为11/2, 提供的噪声容限 $NM_L = 210mV$; 而对于6/6的上拉管, 下拉管版图尺寸为10/2, 提供的噪声容限 $NM_L = 211mV$ 。

以下继续分析功耗最坏条件。固定 $V_{DD} = 5.5V$, 改变器件的栅长及模式等等, 使之进入最坏条件, 就得到10/10上拉管的功耗为0.82mW, 而6/6上拉管的功耗为0.76mW。分析将0.3pF 电容充电到4V所用的时间, 可以得到对10/10的上拉管的最坏条件的延时为29ns, 对6/6上拉管则最坏条件的延时为33ns。用10/10上拉管比用6/6上拉管时功耗延时乘积大约改善5%。

CMOS

最常用的CMOS 最坏条件为“SS”(慢速工作的pMOS FET 和慢速工作的nMOS FET)。像在nMOS 情况一样, 在慢速工作情况下 C_{ox} 是低的, 而所有其它电容都是

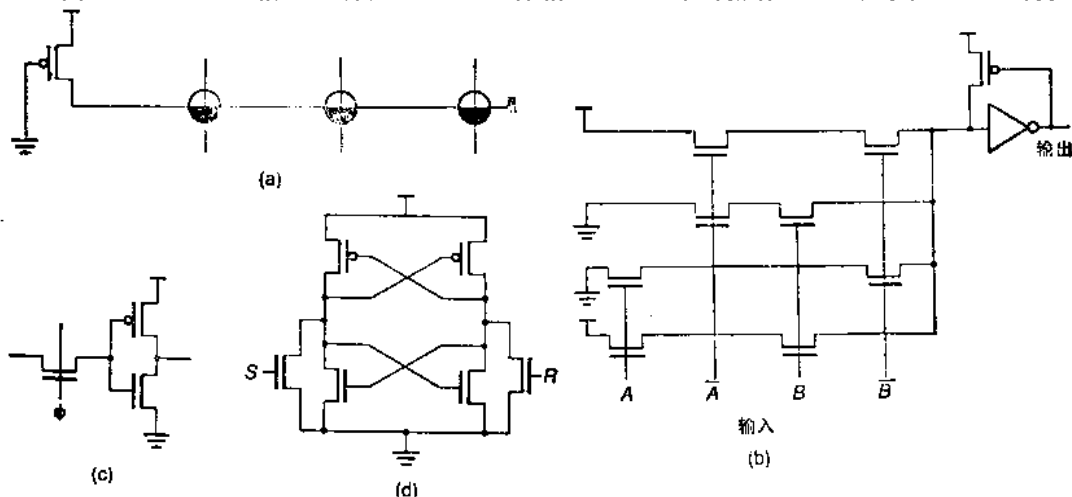


图4.30 最坏条件区域不明显的几种CMOS电路。(a) 有比或非门, (b) 用n沟道选择器的接近nMOS的异或门, (c) 未用互补传输管的1/2移位寄存器单元, (d) 有比S-R锁存器

高的。

在“FF”情况下噪声容限往往是最坏的。这是由于此时器件的阈值电压低而 iR 压降大的缘故。当时钟设计方法中存在竞争条件时，“FF”可成为最坏条件，因为正是在这个工作区域中，对于时钟边沿而言，门的动作是最快的，特别是用多晶硅线提供时钟分配时更是如此（ R_{poly} 很大时会导致时钟畸变的最坏条件）。

图 4.30 中给出了一些有比的 CMOS 电路。这些电路的最坏条件可在“SF”或“FS”条件下发生。

4.5 小 结

噪声和工艺偏差是 VLSI 中实际存在的事实。原则上讲噪声起源于电阻和电容的耦合。尽管存在噪声，数字电路仍旧可以可靠地工作。噪声容限告诉我们怎样去控制噪声，而最坏条件设计则指导我们怎样去控制工艺偏差。最坏条件设计成功的关键在于考虑芯片上各工艺参数之间的相互关系。最坏条件区域是随设计方法而变化的。

我们用反相器研究了基本的电路问题，在这个研究过程中，多次注意到了器件物理界面上的问题。

习 题

- 4.1 有一个理想反相器，其电压传输函数如图 4.31 所示。若 $V_{IL} = 2.5\text{ V}$ 和 $V_{IH} = 3.5\text{ V}$ ，问 V_{OL} 、 V_{OH} 、 NM_L 、 NM_H 、 NR_L 和 NR_H 各为多少？要求 $NM_L + NM_H$ 为最佳时它的噪声容限是多少？

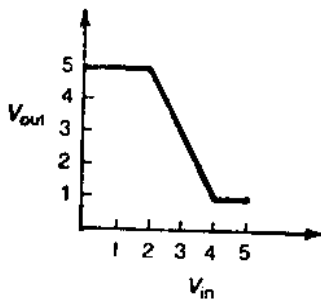


图 4.31

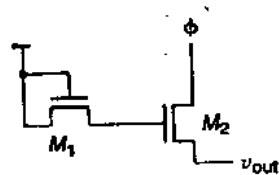


图 4.32

- 4.2 若要求图 4.32 中的电路实现自举，什么是 ΔL_{poly} 、 ΔW 和 T_{ox} 的最坏条件？如果它们中的任何一个达到了相应的条件，它的参量文件是什么？采用附录中给出的 $2\mu\text{m}$ nMOS 工艺画出该电路的版图。为使输出 v_{out} 上升到 V_{DD} ， M_2 的栅宽应该是多少？用 SPICE 程序进行模拟和比较。
- 4.3 解释图 4.33 中电路是如何工作的。这个电路的好处是什么？
- 4.4 对图 4.34 至图 4.36 中所示的三个电路，推导它们与时间有关的输出电压近似表达式。
- 4.5 有一个具有 6/6 上拉管和采用 $2\mu\text{m}$ 工艺的 nMOS 反相器，假定它的输入信号是通过一个 6/2 增强型传输管加来的，问该反相器下拉管的最小栅宽是多少？假设 $NM_H = 600\text{ mV}$ ，其它所有规范均取自示例 4.3。在模拟输入端最大上拉电压时，最好去掉 SPICE 程序中的次开启标志，问最坏条

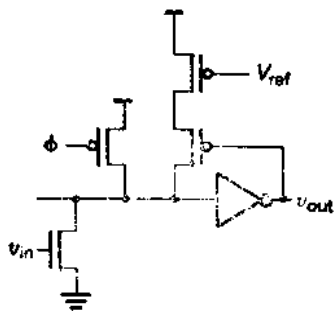


图 4.33

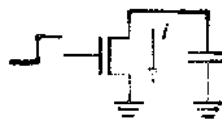


图 4.34

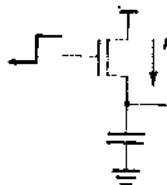


图 4.35

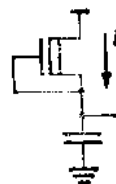


图 4.36

件下 V_{TS} 是高还是低? Δ

- 4.6 在多大输出电压时耗尽型负载 n MOS 反相器的微分增益等于 -1 ? 在这些点处输出电压值是多少?
- 4.7 在没有传输管而采用耗尽型上拉管和增强型下拉管的门电路系列中, 确定它们对 β 比的要求。已知 $V_{DD} = 2.5 \text{ V} \pm 0.2 \text{ V}$, $NM_H = 600 \text{ mV}$ 和 $NM_L = 200 \text{ mV}$ 。问最坏条件下 V_{TS} 是高还是低? Δ
- 4.8 金属作连线是极好的, 因为金属的电迁移限制允许平均电流密度为 $J_{\max}$ 。连线单位长度上的最大平均压降等于 $\rho J_{\max}$ 。这里的 ρ 是金属的电阻率。问对于铝该电压降是多少?
- 4.9 Rideout (Rideout 81) 曾经建议在加工工艺过程中, 在各个点上测量工艺参数, 并且用这些测量结果去控制下一步工艺。问哪些工序影响最大? 一些实际的考虑是什么? 它对最坏条件设计有什么影响? $\Delta \Delta$
- 4.10 电源线和地线的大电容与金属层电阻合在一起形成一个滤波结构。人们想像中它将会有效地滤除电源的噪声。试说明这个滤波器实际上并不十分有效。 Δ
- 4.11 在 CMOS 反相器输入信号开关动作非常慢的时候, 它耗散的能量几乎全部来自超额电流, 其开关能量可以近似为 QV_{DD} 。假定输入电压摆率 $s \text{ (V/s)}$ 为恒定值。试证明在忽略 δ 作用时 Q 可由

$$Q = \frac{K_N S_N}{s(1+\alpha)} (V_{DD} + V_{TP} - V_{TN})^2$$

给出。其中

$$u = \frac{S_N K_N}{S_P K_P} \Delta$$

- 4.12 比较 P 型上拉管和 n 型上拉管的性能。已知 $L = 2 \mu\text{m}$, $T_{OX} = 300 \text{ \AA}$, $\mu_p = 250 \text{ cm}^2/(\text{V} \cdot \text{s})$, $\mu_n = 600 \text{ cm}^2/(\text{V} \cdot \text{s})$, $\gamma = 0.2 \text{ V}^{0.5}$, $V_{TN} = 0.8 \text{ V}$ 和 $V_{TP} = -0.8 \text{ V}$ 。

试确定晶体管的栅宽 W , 使该管能把一个 20 mA 电流源负载加到 3 V 上。假定 $V_{DD} = 5 \text{ V}$ 。若 n 型

管输入电压为5V, P型管输入电压为0V, 试确定哪个管子更宽一些?用手算推导出在两个器件的栅宽相等时输出电压的表达式。 Δ

4.13 在4.4节中, 我们举了一个含有1000个反相器的LSI芯片示例。假定电容变化 $\pm 30\%$, 重新进行分析。

4.14 考虑如下情况: 一个nMOS增强型晶体管的漏、源和衬底都接地, 开始时它的栅极也是低电平。一个电压脉冲加到该管的栅极对它和其它本征电容充电, 这时必然也要从漏、源和衬底提供电荷, 以便平衡加在栅极上的电荷。

我们要研究SPICE是如何模拟这个过程的。采用下列方法去检测充到各引出端的电荷和从各引出端流出的电荷, 而不影响测量条件: 用几个独立的零伏电压源把漏、源和衬底接地。用受电流控制的电流源产生与流过四个电压源成比例的电流, 这四个电压源分别接到器件的四个引出端。如果这些电流被馈送到电容上, 每个电容上所提高的电压数值, 将是流入到相应引出端电荷总量的度量。可以调整电流比例常数和电容尺寸, 以便使所得到的电压值处在合理范围。或许你可以尝试着事先就想好这一点。在没有对地直流通路的情况下, SPICE程序不允许把电容与电流源并联。所以要加入一个与电容并联的大电阻(例如1G Ω)。

利用这个方案, 加一个电压脉冲到栅极上, 看看会发生些什么。看看SPICE程序的工作是否合理?电荷守恒吗?精确吗?当栅电压回到它的初始低电压时会发生什么?在阈值点以上和以下, 在沟道区, SPICE程序分别采用了多大的 X_{PN} 值?

参 考 文 献

- [Bode 45] H. W. Bode, *Network Analysis and Feedback Amplifier Design*, Van Nostrand, New York, 1945. (The gain-bandwidth principle, among other wonderful things.)
- [Burns 64] J. R. Burns, "Switching Response of Complementary Symmetry MOS Transistor Logic Circuits," *RCA Review*, 627—661, 1964.
- [Dang 81] R. L. M. Dang and N. Shigyo, "Coupling Capacitances for Two-Dimensional Wires," *IEEE Trans. Electron Devices Lett.* EDL-2, 196—197, 1981.
- [Demoulin 79] E. Demoulin, J. A. Greenfield, R. W. Dutton, P. K. Chatterjee, and A. F. Tasch, Jr., "Process Statistics of Submicron MOSFETs," *IEEE International Electron Device Meeting*, pp. 34—37, Washington, D. C., 1979.
- [Fang 75] F. F. Fang and H. S. Rupprecht, "High Performance MOS Integrated Circuits Using Ion-implantation Techniques," *IEEE J. Solid-State Circuits* SC-10, 205—211, 1975.
- [Forbes 73] L. Forbes, "n-Channel Ion-implanted Enhancement/Depletion FET Circuit and Fabrication Technology," *IEEE International Solid-State Circuits Conf.* SC-8, 226—230, New York, 1973.
- [Glasser 81] L. A. Glasser, "The Analog Behavior of Digital Integrated Circuits," *Proceedings of the Eighteenth Design Automation Conference*, pp. 603—612, Nashville, Tenn., 1981.
- [Haviland 80] G. L. Haviland and A. A. Tuszynski, "A CORDIC Arithmetic Processor Chip," *IEEE J. Solid-State Circuits* SC-15, 4—15, 1980.
- [Hill 68] C. F. Hill, "Noise Margin and Noise Immunity in Logic Circuits," *Microelectronics* 1, 16—21, 1968.
- [Lo 61] A. W. Lo, "Some Thoughts on Digital Components and Circuit Techniques," *IRE Trans. Electronic Computers* EC-10, 416—425, 1961. (Includes discussion of nonreciprocal logic elements.)
- [Lohstroh 79] J. Lohstroh, "Static and Dynamic Noise Margins of Logic Circuits," *IEEE J. Solid-State Circuits* SC-14, 591—598, 1979.
- [Lohstroh 83] J. Lohstroh, E. Seevinck, and J. De Groot, "Worst-Case Static Noise Margin Criteria

- for Logic Circuits and Their Mathematical Equivalence," *IEEE J. Solid-State Circuits* SC-18, 803—807, 1983.
- (Maly 82) W. Maly and A. J. Strojwas, "Statistical Simulation of the IC Manufacturing Process," *IEEE Trans. Computer-Aided Design* CAD-1, 120—130, 1982.
- (Masuhara 72) T. Masuhara, M. Nagata, and N. Hashimoto, "A High-Performance n-Channel MOS LSI Using Depletion-Type Load Elements" *IEEE J. Solid-State Circuits* SC-7, 224—231, 1972.
- (May 79) T. C. May, "Soft Errors in VLSI, Present and Future," *IEEE Trans. Components, Hybrids, Manufacturing Technology* CHMT-2, 377—387, 1979.
- (Polkinghorn 70) R. W. Polkinghorn et al., "FET Driver Using Capacitor Feedback," U. S. Patent 3,506,851, 1970.
- (Rideout 81) V. L. Rideout, "Trends in Silicon Processing," *Proc. Second Caltech Conf. Very Large Scale Integration*, C. L. Seitz (ed.), pp. 65—110, Pasadena, Calif., 1981.
- (Sai-Halasy 83) G. A. Sai-Halasy, "Cosmic Ray Induced Soft Error Rate in VLSI Circuits," *IEEE Electron Device Lett.* EDL-4, 172—174, 1983.
- (Stein 77) K.-U. Stein, "Noise-Induced Error Rate as Limiting Factor for Energy per Operation in Digital IC's," *IEEE J. Solid-State Circuits* SC-12, 527—530, 1977.
- (Stone 84) D. C. Stone, J. E. Schroeder, R. Kaplan, A. R. Smith, "Analog CMOS Building Blocks for Custom and Semicustom Applications," *IEEE Trans. Electron Devices* ED-31, 189—195, 1984 (Noise sources in analog circuits.)
- (Tokuda 83) T. Tokuda, K. Okazaki, K. Sakashita, I. Ohkura, and T. Enomoto, "Delay-Time Modeling for ED MOS Logic LSI," *IEEE Trans. Computer-Aided Design* CAD-2, 129—134, 1983.

第五章 电路技术

在第二至第四章中, 我们学习了为使 VLSI 具有一定的功能在设计时必须考虑的因素。同时也研究了 MOS 结构的物理和模型建立, 研究了逻辑抽象与工艺技术所带来的限制以及产品成本和功耗问题。除这些基本内容之外, 我们还打算了解一下电路形式以及解决特殊电路问题的惯用方法。大多数这些特殊电路技术是作为基本电路问题的范例的, 但其意义并不限于某一个特殊电路。在本章, 我们将研究一批 MOS 数字电路, 同时还要介绍学习这些电路所必需的概念和分析工具。这些电路提供了解决电路技术问题的原始资料并可作为创新和进一步学习的基础。

5.1 简单逻辑门的电路优化技术

我们研究的第一个问题是简单逻辑门的优化。VLSI 需要优化的主要变量是晶体管栅宽, 在这一节中, 我们将分析关键通路 [Glasser 84] 中晶体管尺寸的优化问题, 并由此导出一些公式和不等式, 它们可以对更复杂的情况提供一般性指导。在这种分析中采用在 4.2 节中所导出的 MOS 逻辑门 $R_i C_i$ 模型。通过逻辑门组成的链路的总延时 T 可以写成各个 RC 之和的形式。可有

$$T = \sum_{i=0}^N t_i \quad (5.1)$$

其中

$$t_i = R_i \left(C_{i+1} + Y_i + \frac{\alpha_i C_i}{n_i} \right) n_i \quad (5.2)$$

这里的 R_i 是第 i 级下拉管的等效电阻, C_{i+1} 是第 $i+1$ 级的有效输入电容, Y_i 是连接第 i 级和 $i+1$ 级连线的总等效连线电容, Y_i 也包括不在该关键通路中的其它各级的等效输入电容。第 i 级的固有电容负载由 α_i 表征。当晶体管的栅宽增加时, 它的寄生参数也增加。作为一级近似, 这些增加都归在 α 项中, 这里的 α_i 往往是常数 (即对于所有 i , $\alpha_i = \alpha_0$), 图 5.1 说明了这些参数。参数 n_i 称为扇出数, 根据求解问题的不同, 它有许多不同的解释, 其基本目的是说明下拉网络之间的区别。图 5.2 是一个简单的逻辑门链路。在这个示例中, 可有

$$T = R_0 \frac{\tau}{R_1} + R_1 \left(Y_1 + \frac{\tau}{R_2} \right) + 2 \beta_r R_2 \left(Y_2 + \frac{\tau}{R_3} \right) + R_3 C_{load} \quad (5.3)$$

或

$$T = 5R_0 C_r + \frac{R_r}{5} (Y_1 + 10C_r) + \frac{\beta_r R_r}{5} (Y_2 + 10C_r) + \frac{R_r}{10} C_{load} \quad (5.4)$$

利用式 (4.36), 可以把式 (5.1) 重写为

$$T = \tau \sum_{i=0}^N \left[R_i \left(\frac{1}{R_{i+1}} + \frac{r_i}{\tau} \right) + \frac{a_i}{n_i} \right] n_i \quad (5.5)$$

其中 $n_0 \equiv 1$, R_0 是源区等效电阻, $C_{N+1} = 0$. 在式 (5.5) 中, 我们规定了 $Y_N \equiv 0$ 和 $R_{N+1} \equiv \tau / C_{load}$, 这里的 C_{load} 是输出电容, 这样做纯粹是为了便于用符号表示。

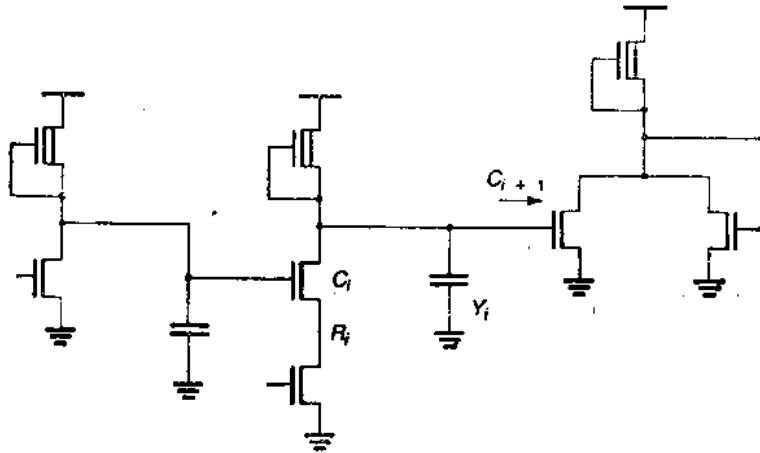


图5.1 一组逻辑门中的重要电容和电阻, $\tau = R_i C_i$

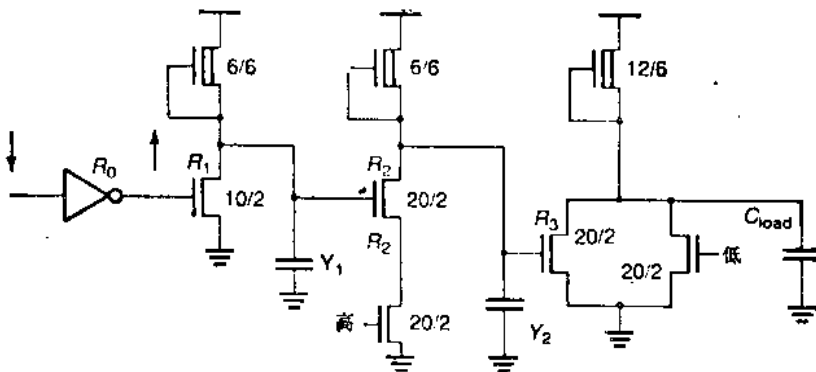


图5.2 说明 $\tau = RC$ 延时模型的元件示例

式 (5.5) 是本节中最重要的公式。我们首先以闭式对一些特殊情况进行求解。最重要的特殊情况是所有连线电容 Y_i 远远小于相应的输入电容 $C_{i+1} = \tau / R_{i+1}$ (除 C_{load} 外), 在最小延时的条件下, 可能出现这种情况, 这时驱动管的栅非常宽。通过求出 T 对 R_i 的梯度并令它等于 0, 可以找出最佳速度的条件:

$$0 = \nabla_{R_i} T = -\frac{R_{i-1} n_{i-1}}{R_i^2} + \frac{n_i}{R_{i+1}} \quad (\text{对于 } i = 1 \sim N) \quad (5.6)$$

在此情况下我们有

$$t_i = n_i \tau \left(\frac{R_i}{R_{i+1}} + \frac{a_i}{n_i} \right) \quad (5.7)$$

可以看出, 在 $a_i = a_0$ 的特殊情况下, 从式 (5.6) 可以得出所有 t_i 相等, 即

$$t_i = t_{i+1} \equiv t_0 \quad (5.8)$$

在此特殊情况下，每一级的延时都相等，而这与有些门是否是或非门以及其它门是否是或非门无关。定义 t_0 如下：

$$t_{00} \equiv t_0 - \alpha_0 \tau \quad (5.9)$$

式 (5.6) 可以用闭式求解，其结果是

$$R_i = \tau - \frac{n_i - 1}{t_0} R_{i-1} \quad (5.10)$$

$$R_i = R_0 \left(\frac{t_{00}}{\tau} \right)^{i-1} \prod_{j=1}^{i-1} n_j \quad (5.11)$$

$$t_{00} = \tau \left(\frac{R_0}{R_{N+1}} \prod_{j=1}^N n_j \right)^{\frac{1}{N+1}} \quad (5.12)$$

和

$$T_{\min} = (N+1)t_0 \quad (5.13)$$

用这些公式可以对简单门链路的器件尺寸和延迟进行优化的第一步估计。对于所有 $n_i = 1$ 的简单情况，上述结果化简为常见的器件尺寸按比例变化，这时，每个器件的尺寸都是前级器件的尺寸乘以常数。

在不能忽略连线电容的情况下，或者在把功耗或面积看作附加度量标准的情况下，就找不到闭式的解答。我们把总功耗或面积写成如下形式：

$$B = \sum_{i=1}^N b_i \quad (5.14)$$

这就是说，我们把每一级同某一个表征该级所用资源的量联系在一起。在考虑功耗时，我们把 b_i 解释为各级的平均功耗。对于有源区面积，我们则把 b_i 解释为各级的有源区面积。在功耗和有源区面积这两种情况下， b_i 取如下形式：

$$b_i = m_i c_i \quad (5.15)$$

其中 m_i 是取决于逻辑门拓扑关系的一个常数。例如，CMOS 逻辑门的有源区面积正比于逻辑门中晶体管数与各个晶体管面积的乘积。简单地讲，nMOS 逻辑门中的功耗决定于下拉管的栅宽（它直接正比于电容 C_i ）乘以 β 比再乘以与之串联的下拉管个数。因此可有

$$B = \sum_{i=1}^N m_i C_i = \tau \sum_{i=1}^N \frac{m_i}{R_i} \quad (5.16)$$

已知 T ，我们要求 B 最小。这就是说，给定一个延时规范，要求整个电路的面积或功耗最小。为了做到这一点，我们引用边际效用论 (marginal utility argument)。它表明为了使 B （这个从整体讲是稀有的“资源”）最小，就要求从局部看每一级都必须充分利用这些稀有的“资源”。在总延时对每一级所用的总体“资源”的灵敏度都是一样的条件下，这一定是对的。为了弄清楚这个论点，设想总延时对于第五级所有功耗的灵敏度大于对第七级功耗的灵敏度，从第七级减去一些功耗，该电路就会稍微变慢一些，再把该功耗的一部分加到第五级上以便提高它的速度，用以补偿第七级的速度下降。这样处理后，这个电路就将以原来的速度工作但功耗较小，所以是有好处的。这个边际效用论相当于该问题的拉格朗日乘法公式，我们有

$$0 = \nabla_{\mathbf{b}_i} B + \lambda \nabla_{\mathbf{b}_i} T \quad (5.17)$$

它可化简为

$$-\lambda^{-1} = \nabla_{\mathbf{b}_i} T \quad (5.18)$$

求解上式可得

$$\lambda^{-1} = \frac{R_i}{\tau m_i} \left(\frac{n_i R_i}{R_{i+1}} + \frac{n_i R_i Y_i}{\tau} - \frac{n_{i-1} R_{i-1}}{R_i} \right) \quad \text{对于 } i = 1 \sim N \quad (5.19)$$

其中 λ 是拉格朗日乘数。

我们可以作出许多有用评定, 通过设定 $\lambda^{-1} = 0$, 在最佳速度的特殊情况下, 式 (5.19) 简化为式 (5.6)。延时对面积或功耗导数的负值可用 λ^{-1} 表示, 因负的 λ 意味着正处于可以同时降低面积和减少延时的区域, 所以最佳解常常出现在 λ 为非负值的区域。如果碰巧处在这样合适的区域中, 就很难进行优化了。利用 t_i 可以把式 (5.19) 重新写为

$$t_i = t_{i-1} + \frac{\tau^2 m_i}{\lambda R_i} - n_{i-1} R_{i-1} Y_{i-1} + \tau (\alpha_{i-1} - \alpha_i) \quad (5.20)$$

其中, τ , n_i , m_i , R_i , Y_i 和 λ 都是非负值。当 α 为常数时, 这个式子告诉我们, 对于存在连线寄生参数时的最佳速度 ($\lambda^{-1} = 0$), 级延时是其在链中位置的单调下降函数, 这就是说最大延时出现在第一级。从物理上说, 这是因为门链路的前级驱动能力是有限的, 而接近链路末级时, 由于驱动能力依次扩大, 就可能具有更高的驱动水平, 因而就更能够忽略连线电容。

另一方面, 如果没有连线电容, 而功耗或面积又是重要的, 则级延时就是位置的单调上升函数, 通过研究多级驱动大负载的问题就可以弄明白这一点。如果 λ^{-1} 等于 0, 每一级的延时全是一样, 而且延时对于功耗的灵敏度也会是一样的。尽管如此, 最末级功耗所占的百分比仍是最大的。所以, 如果在设计中要求大量减少功耗, 并使芯片运行得慢一些, 则从最后一些级减少功耗将最为有利, 这是因为这些级是消耗大量功率的唯一地方。

最后我们重申, 在 VLSI 设计中, 多条道路中哪一条是关键通路往往是不明显的。况且该通路被优化后, 另外一条通路就会成为关键通路。因为一条通路的优化会影响另一条通路的速度, 如果试图对最关键通路优化进行直接计算, 那么第二条通路就会转化为新的最关键通路, 如此等等, 就会出现大量重复计算。在此提供的技术可以扩展到多通路问题。Matson [Matson 83] 研究了多通路问题的求解, 其中包括了更精确的宏模型。

5.2 大扇出电路和时钟驱动器

在 CMOS 或耗尽型负载 nMOS 工艺中, 当一个逻辑门驱动另一个邻近逻辑门时, 为了优化或改进延时, 除了调整晶体管尺寸和处理版图之外, 没有多少别的措施可采取。反之, 当驱动管尺寸与负载极不匹配时, 就有充分的可能和必要去进行电路的改进和创新。这种情况是常见的, 因为大扇出条件和规则结构是密切相关的。大扇出情况是指一个小的晶体管必须驱动一个非常大的电容。

多级缓冲器 (staged buffers) 和中继器 (repeaters)

当电容负载远大于驱动管的栅电容时, 在驱动器件和负载之间加几级缓冲器往往是很恰当的。每一个缓冲器都比前一级略大一些 [Lin 75], 如图 5.3 所示。假定有 M 级缓冲器, 已知源电阻为 R_0 , 负载电容为 C_L , 则有

$$T = \sum_{i=0}^M R_i C_{i+1} = \tau \sum_{i=0}^M \frac{C_{i+1}}{C_i} \quad (5.21)$$

其中 $C_{M+1} = C_L$ 。从 5.1 节我们知道, 最小延时发生在 $t_i = t_{i+1} = t_0$ 的时候, 因此存在某个常数 g 满足

$$C_{i+1} = g C_i \quad (5.22)$$

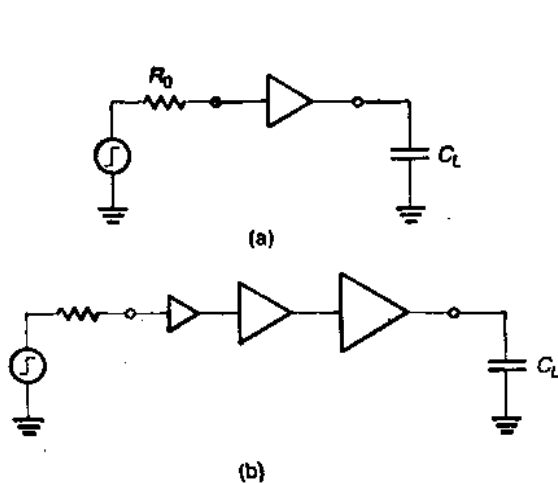


图 5.3 (a) 单级缓冲器;
(b) 变换后的多级缓冲器

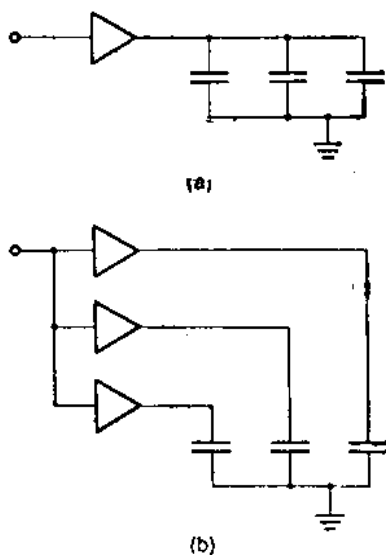


图 5.4 (a) 单级缓冲器; (b) 对不同的关键通路, 单级缓冲器变换为不同尺寸的缓冲器

我们有

$$T = \tau (M + 1) g \quad (5.23)$$

和

$$C_L = g^{M+1} C_0 \quad (5.24)$$

其中, $C_0 \equiv \tau / R_0$ 。合并式 (5.22) 和 (5.24), 可以得到

$$T = \tau (M + 1) \left(\frac{C_L}{C_0} \right)^{\frac{1}{M+1}} \quad (5.25)$$

对 M 进行优化, 我们得到

$$g = e \quad (5.26)$$

和

$$M + 1 = \ln \frac{C_L}{C_0} \quad (5.27)$$

最小总延时是

$$T = e \tau \ln \frac{C_L}{C_0} \quad (5.28)$$

值得注意的是最佳级数和总延时只随驱动与负载间的失配比按对数关系增加。在包含多个负载时，可以把它们分开并单独优化每一条通路。利用这种技术，可以避免因驱动关键通路中的负载而消耗功率。这种技术示于图 5.4 中。

在转到更复杂的讨论之前，还可以从简单的电阻模型中取得最后一点直观知识。让我们研究用中继器断开长电阻线以改进延时的问题。我们知道，随着连线的增长，其延时按平方律上升。如果插入中继器，尽管中继器在连线中引入了某些附加的延时，但仍旧改善了信号的上升时间。假如把一条 RC 长线断开为 $2N$ 段，线的总电阻为 R_L ，总电容为 C_L ，则每一段具有的电阻为 $R_L/2N$ ，电容为 $C_L/2N$ 。我们定义 τ_L 为：

$$\tau_L = \frac{R_L C_L}{2} \quad (5.29)$$

这里的 τ_L 为未激励线的“时间常数”¹⁾。如图 5.5 所示，优化线的每一段都由一个 MOS 反相器开始。

假定采用 n MOS 耗尽型负载工艺技术。每一段 RC 线的末端都接一个负载电容 SC_L ， S 为下拉管的形状因子，零级上拉管电阻和下拉管电阻分别为 $\beta_u R_i/S$ 和 R_i/S 。估计反相器对性能的最简单方法是把电容和电阻相加，把每个电阻乘以其右侧全部电容

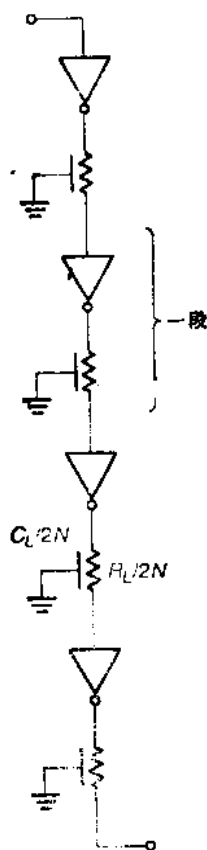


图 5.5 带有中继器的 RC 线模型

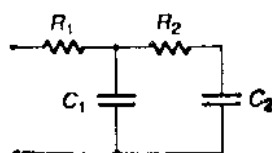


图 5.6 在双时间常数系统中延时的估计。 $T = R_1(C_1 + C_2) + R_2 C_2$

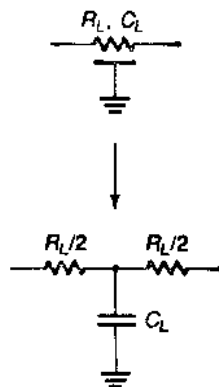


图 5.7 用 π 模型表示 RC 线的模型

1) 从 2.4 节我们知道了估算延时的较好方法，然而要记住这些推导都是近似的结论。

之和, 如图 5.6 所示。RC 线用如图 5.7 所示的 T 网路模拟。可有

$$t_{\text{pair}} = \left[\frac{(1 + \beta_r) R_r}{S} + \frac{R_L}{2N} \right] \left(SC_r + \frac{C_L}{2N} \right) + \frac{R_L C_L}{4N^2} \quad (5.30)$$

该链路的总延时为

$$T = N t_{\text{pair}} \quad (5.31)$$

取它的梯度等于 0, 可以使 T 最小, 则有

$$0 = \nabla_{s,N} T \quad (5.32)$$

该式可简化成下列联立方程,

$$\begin{cases} 0 = (1 + \beta_r) R_r C_r - \frac{R_L C_L}{2N^2} \\ 0 = \frac{R_L C_r}{2} - \frac{(1 + \beta_r) R_r C_L}{2S^2} \end{cases} \quad (5.33)$$

$$(5.34)$$

代入等式 $\tau_L = R_L C_L / 2$ 和 $\tau = R_r C_r$, 我们得到

$$N = \sqrt{\frac{\tau_L}{2(1 + \beta_r)\tau}} \quad (5.35)$$

$$S = \sqrt{(1 + \beta_r) \frac{R_r}{R_L} \frac{C_L}{C_r}} \quad (5.36)$$

求解最佳延时, 得到

$$T_{\text{opt}} = (1 + \sqrt{2}) \sqrt{1 + \beta_r} \sqrt{\tau \tau_L} \quad (5.37)$$

最佳延时是某个比例常数乘以本征速度参数 τ 与连线速度参数 τ_L 的几何平均值。随着工艺技术的改进, τ 也会得到改善, 但是 τ_L 往往不变。从式 (5.35) 和 (5.37) 可以得出结论: 在先进的技术中, 中继器的采用将会变得更加普遍。

超缓冲器 (Superbuffers)

为了驱动大电容负载要耗费大量功率和芯片面积。当负载非常大时, 要采用特殊的电路技术以求获得更有效的 (或许更为复杂的) 解决办法。在 nMOS 电路中, 静态功耗非常大, 对此, 引用了两种基本技术。其一, 为了节约功耗采用了推挽输出级, 其目的是把 nMOS 作成类似于 CMOS 的电路, 在稳态时, 上拉管和下拉管不同时导通。第二个基本技术是使晶体管过驱动¹⁾, 以便在开关转换期间, 每微米晶体管栅宽流过的电流大于正常情况。

图 5.8 给出了一般化的超缓冲器和一些可能的构造。其中电路形式 (b) 具有两个耗尽型上拉管和两个增强型下拉管, 这个电路最易于分析, 所以我们将用它作为示例。对于其它电路形式, 其细节肯定有差别, 但它们的基本折衷方案是类似的。

举一个我们可以分析处理的例子。假定 $V_{TD} = -V_{DD}/2$ 且忽略体效应。这个例子中还忽略了前置驱动级中的延时和功耗。我们考查一下该缓冲器将负载电容 C 从地电平充电到 $V_{DD}/2$ 所用的时间。假定输入电压是从 V_{DD} 到 0 的阶跃电压。对于标准型式的耗尽型负载反相器, 负载管处在饱和状态。流过上拉管的电流是

$$I_{pu} = K' S V_{TD}^2 \quad (5.38)$$

1) 在高度优化的技术中, 这些过驱动电压会引起可靠性问题。在 CMOS 电路中, 自举电压可引起自锁效应。

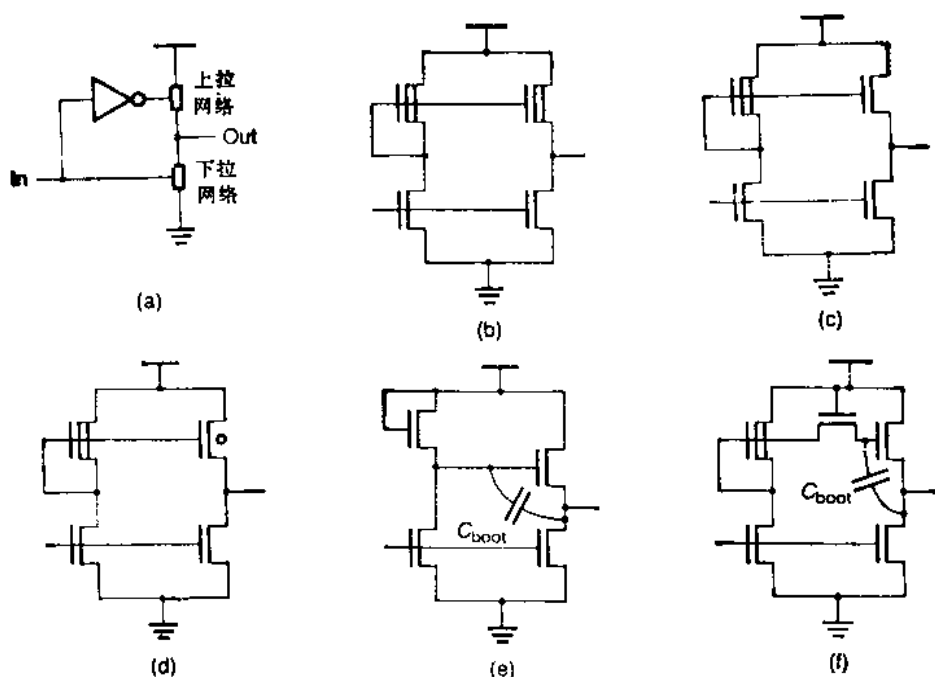


图5.8 各种超缓冲器

在延迟时间 t_D 内, 输出电压以斜波形式上升到 $V_{DD}/2$ 。 t_D 由下式给出,

$$t_D = \frac{C}{-2K'SV_{TD}} \left(\frac{V_{DD}}{-V_{TD}} \right) \quad (5.39)$$

在超缓冲器中, 有两种可能情况。如果是小负载电容, 则输出级响应比预充级快的多。在这种情况下, 耗尽型驱动管的 V_{GS} 电压近似为 0, 除了它的固有电容和占用面积略大一些以外, 该超缓冲器基本上等效为一个反相器。反之, 如果是大负载电容 C , 预充级响应远远快于驱动输出级。在极端情况下, 上拉管的 V_{GD} 瞬间达到 0 伏。因为这时上拉管处在非饱和区。可有

$$I_{pu} = K'S(V_{DS} - 2V_{TD})V_{DS} \quad (5.40)$$

对上拉管电流进行积分, 我们有

$$t_D = \frac{C}{K'S} \int_{V_{DD}}^{V_{DD}/2} \frac{dV_{DS}}{V_{DS}(V_{DS} - 2V_{TD})} \quad (5.41)$$

求解上式得到

$$t_D = \frac{C}{-2K'SV_{TD}} \ln \frac{V_{DD} - 4V_{TD}}{V_{DD} - 2V_{TD}} \quad (5.42)$$

引用 $V_{TD} = -V_{DD}/2$, 可得出该超缓冲器比同样尺寸的耗尽型负载反相器快 5 倍。由于稳态条件下二者的负载电流是一样的, 所以这个结果意味着超缓冲器的功耗延迟乘积也大约改善了 5 倍。如果调整超缓冲器的输入电容和直流功耗分别与相应反相器的输入电容和直流功耗相等, 这个比较就更加精确。图 5.9 中画出了这两种电路的版图。图 5.10 中为上升延迟时间 t_D (输出端上升到 $V_{DD}/2$) 与负载电容 C 的关系, 这些模拟结果包括了版图电容。在这个示例中, 超缓冲器的版图约增大 10%, 而随着缓冲器的加大, 这个百分比还要下降。最小尺寸的超缓冲器比最小尺寸反相器可大出 50%。

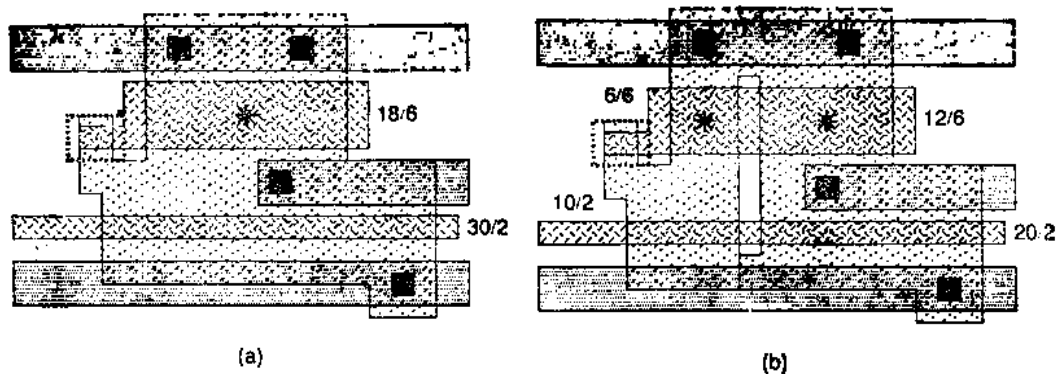


图5.9 具有等效输入电容和直流功耗的两种电路。(a) 反相器；(b) 超缓冲器

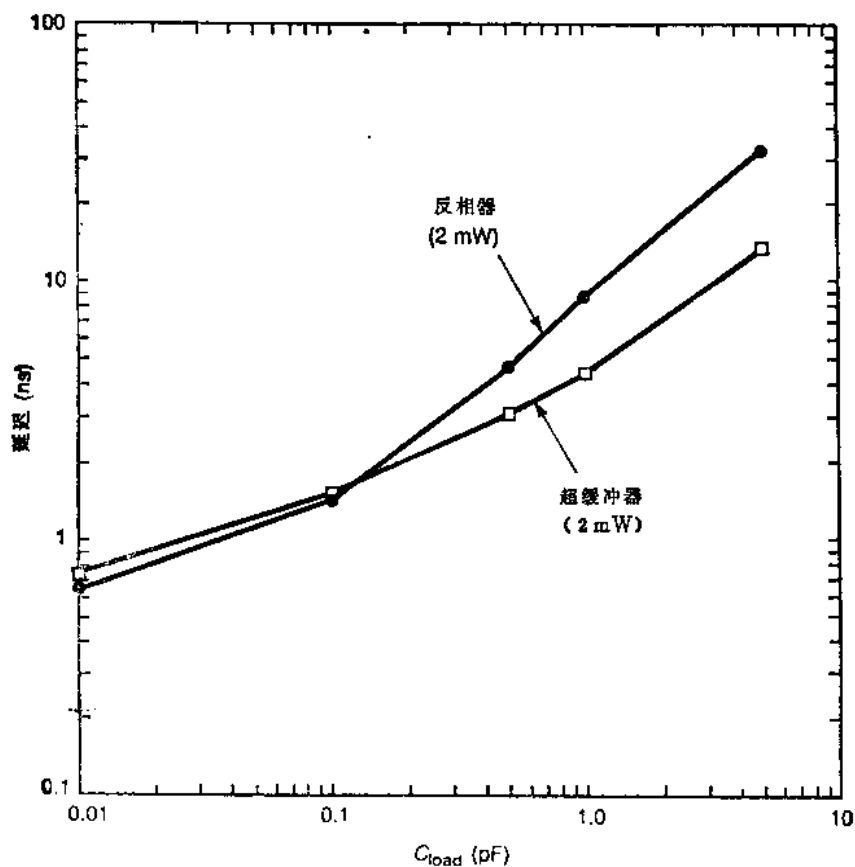


图5.10 图 5.9所示的超缓冲器和反相器的延时与负载电容的函数关系

超缓冲器的耗尽型驱动管有一个极好的特性，它有可能使输入电容和功耗保持常数，而在其它的缓冲器设计中，这些参数相互关系不密切。例如，对于同样的输入电容，图 5.8 (d) 所示电路的功耗较小，但是比图 5.8 (b) 中的电路慢一些。这就不便于对它们进行比较。

回顾在 4.2 节中对各种负载线 $I-V$ 特性的讨论可知，图 5.8 中所示的超缓冲器具

有与时间相关的负载线。图 5.11 表示出其中两种负载线的特征。

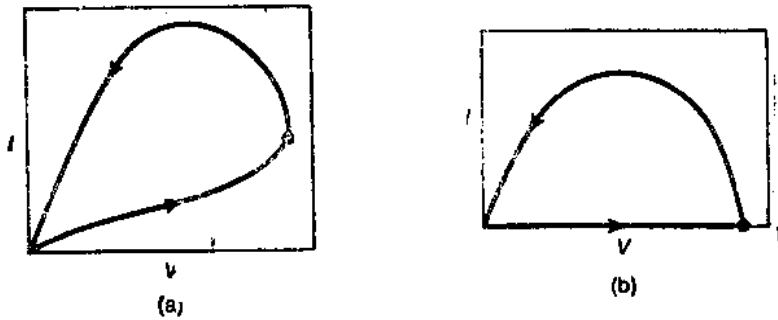


图5.11 超缓冲器与时间相关的负载线。(a) 对应于图 5.8(b)，(b) 对应于图 5.8(d)

时钟缓冲器 (clocked buffers)

地和 V_{DD} 并不是同步 MOS 芯片上的唯一低阻抗功率源。不论是外接时钟还是精细产生的内部时钟，都是极好的动态功率源。在大规模 TTL 电路设计中，常常坚持规定不用时钟作为选通脉冲，而在 MOS VLSI 中并不是这样，考虑一个普通情况：在这种情况下，一个时钟相结束时，控制信号变为有效，然后，在下一个时钟相开始时，控制信号再度有效，并且，必须驱动许多负载。我们可以使用驱动器。然而，还有另外一种几乎不占用功耗和面积，而且对于时钟完全同步的技术。

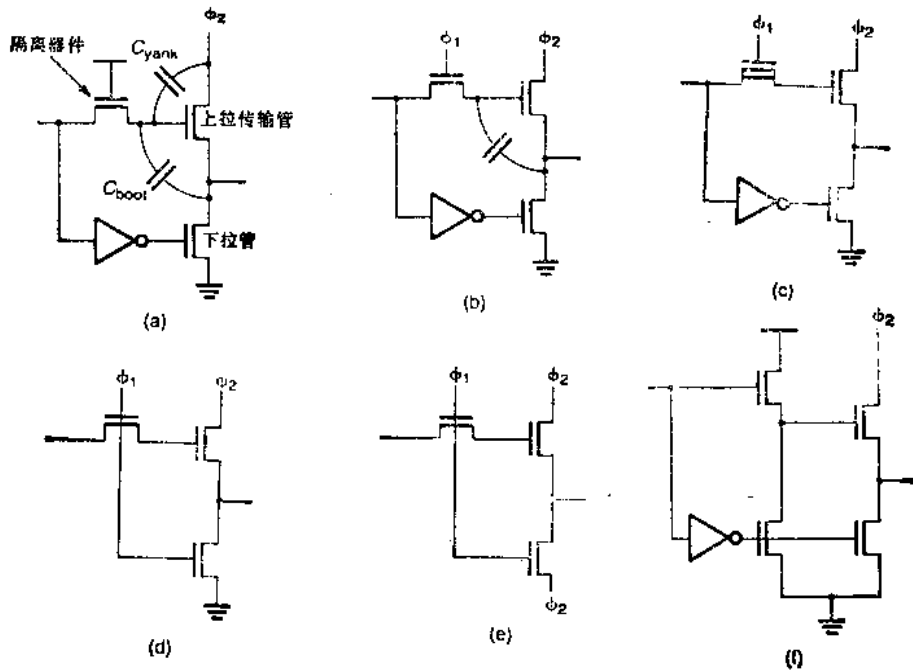


图5.12 各种时钟缓冲器的驱动。这些电路的区别在于定时、低阻抗输出、所要求的时钟数目和输入必须保持有效的时长短

图 5.12 示出了时钟缓冲器的各种变型，其共同点是所有这些设计都包含三个元件。门时钟用上拉传输管与输出端分离。这种传输管的栅往往是较宽的，同时被过驱动以便

使时钟与输出之间的失真最小。输出电压摆幅是从地变到最正的电压（典型值是 V_{DD} ），因而该电路具有强低电平和强高电平输出。该电路的第二个元件是一个隔离網路，它把输入信号耦合到上拉传输管的栅极上，使该栅节点能被举到大于 V_{DD} 的电压。第三个元件是一个下拉管，它保证在输出低电平时输出端始终为地。

这些电路的定时细节有些不同。例如，参看图5.12，电路（a），（b），（c）和（f）均要求在整个 ϕ_2 期间输入端都保持高电平，否则上拉传输管将会关断，输出端将被拉成低电平。在电路（d）和（e）中，在 ϕ_1 变为低电平后，输入向低电平转换就不会有什么影响。然而在电路（b）中，在这种情况下就会有竞争。这些电路中的一部分，例如图（e），对版图是非常敏感的，这是因为这些电路对版图电容的高灵敏度所致。电路（c）采用动态耗尽技术，在那里耗尽管实际上是关断的¹⁾。该电路的好处是可以提供较高的自举电压。

图5.12（a）中有两个耦合电容，分别记为 C_{boot} 和 C_{yank} ，它们可以产生过驱动电压。 C_{yank} 从 C_{GDO} 变化到 $C_{GDO} + WL C_{OX} / 2$ ， C_{boot} 从 $C_{GSO} + WL C_{OX} / 2$ 变化到 $C_{GSO} + 2 WL C_{OX} / 3$ 。为了使用 C_{yank} ，就得利用变容二极管效应。当晶体管的栅电压超过阈值电压时，栅源电容是大的，但是当 $V_{GS} < V_T$ 时，电容是小的。因而 C_{yank} 和寄生参量必须被设计成只在 $V_G > V_T$ 时而不是在 $V_G < V_T$ 时栅电压才被猛拉为高电平，这是一个速度非常快的电路。自举电容的机构并不复杂。当输出上升时，输出电平将栅节点电平举起来。如果传输上拉管完全关断，输出就不再上升，从而不产生自举效应。

电路处在有大量耦合噪声的环境中时，电路（d）和（e）可能存在问题。在 ϕ_2 期间，输出高电压具有低阻抗，而输出低电压却具有高阻抗。另一方面，这些设计是比较简单的，在输出线上加一个小功率锁存器就可以抗住噪声脉冲而保持它为低电平。这个锁存器可以加到远离驱动器输出线的另一侧。这样就加大了版图设计的自由度。图5.13是一个典型的结构。

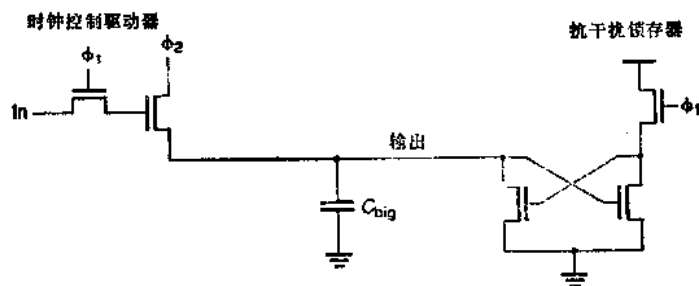


图5.13 具有抗噪声保持低电平锁存器的时钟缓冲器。

时钟驱动器

与缓冲器不同，在时钟驱动器的设计中，往往有可能用上升时间换取延迟时间。对缓冲器来说，其任务是尽可能快地把数据从一个地方送到另一个地方，而在许多时钟应用中（当缓冲外来时钟时，我们所讨论的技术就不太适用），时钟边沿的质量则是极端

1) 第二章中曾介绍过，由于注入杂质较深而使耗尽型器件难以夹断。在 $V_G = V_{DD} - V_{PB}$ 的情况下耗尽型器件夹断不断的可能性是极小的。尽管如此，有关这种应用的新技术，在采用以前应加以说明。

重要的。产生陡峭时钟边沿的最好途径是采用“把它卷起来并突然折断”的方法。

我们将研究五种不同的 nMOS 时钟驱动器，其中每种驱动器都采用了越来越完善的和复杂的电路，这些只是该类电路许多可能变型中的一小部分。图 5.14 为第 I 种情况，这是一种最简单的驱动器。该驱动器驱动输出直到高达 V_{DD} 和低到地电平，对于时钟来说这一点是重要的特性。该驱动器驱动 20 pF 的负载。根据工艺技术和电路型式不同，时钟负载的变化范围可以从低于 10 pF 到超过 100 pF。例如，当时钟必须驱动一些晶体管的源区、漏区和它们的栅极时，负载就会更大一些。在情况 I 中，晶体管 M_3 的作用是当 v_2 自举时与 v_1 隔离。 M_4 管的栅足够宽，因而不需要外部附加自举电容。这个电路正确工作的关键是 v_{out} 不能很快地跟随 v_2 （如果 v_{out} 确实很快地跟随 v_2 ， v_2 就不能超过 V_{DD} ）。图 5.15 给出该电路的 SPICE 模拟结果，模拟时采用了典型参数。因为 v_{out} 升高，所以要在 v_2 上叠加所预期的高自举电压是困难的。

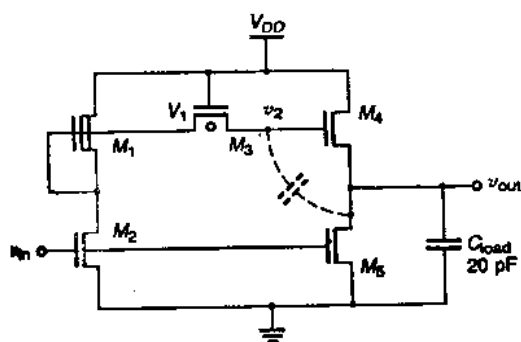


图 5.14 第 I 种情况：自举驱动器

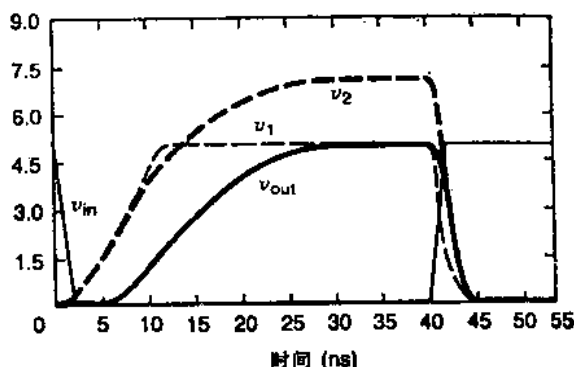


图 5.15 第 I 种情况的电压波形

图 5.16 的电路为第 II 种情况，它通过使 M_5 栅极电压的下降延迟解决了自举问题 [Schroeder 75]，这样就可以在较长时间内保持输出电压为低电平，能够在 v_{out} 开始上升之前使 v_2 充到较高的电压。图 5.17 说明了这第 II 种电路的模拟响应结果。作为一级近似，可以用上升时间替换延时，这是因为该电路延迟了自举电容的充电，而自举电容充电越快输出的转变也就越快。延时是由晶体管 M_5 至 M_{10} 提供的。

图 5.18 中的第 III 种情况说明了控制自举所需隔离晶体管的另一种不同方法。在这种情况下，在自举之前， M_3 可以作为一个被驱动的耗尽管使 v_2 充到 V_{DD} 。这一点对于不能采用零阈值的晶体管，或零阈值晶体管具有大的体效应系数时特别重要。这个电路的性能与第 II 种电路相似。

在情况 I 至情况 III 中，自举作用是以负载管的工作速度完成的。在图 5.19 所示的情况 IV 中，采用单独的输出级 (M_{11} 和 M_{12}) 来驱动一个附加的自举电容 M_{13} [Proebsting 75]。这就使自举电容上的电压 V_0 能够比上述情况更快地卡断。图 5.20 表示情况 IV 的电压波形。

图 5.21 比较了四种情况下的输出电压。该图清楚地表明了上升时间对延时的折衷关系。由于带有延迟线的电路对于加在其输入端上的假信号很敏感，所以在驱动这些电路时必须小心从事。

有人可能要问，为什么需要等信号来到时才开始对自举电容充电呢？这是由于电路

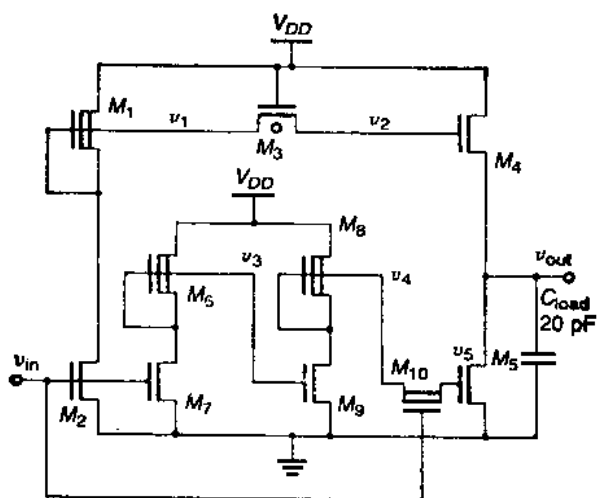


图5.16 第Ⅰ种情况：具有延时自举的自举超驱动器

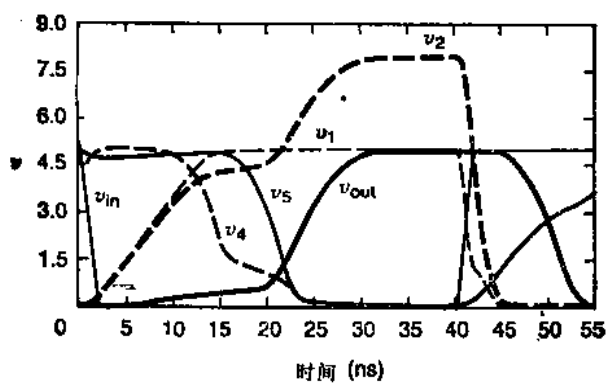


图5.17 第Ⅰ种情况的电压波形

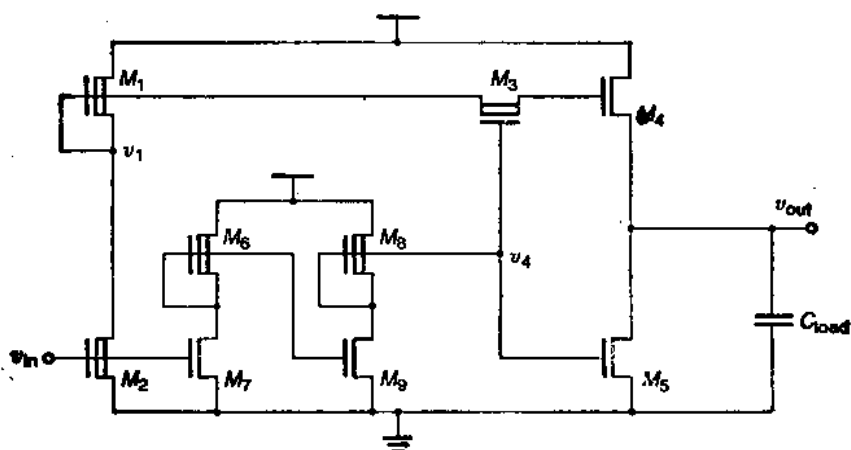


图5.18 第Ⅱ种情况：具有动态耗尽隔离管的自举超驱动器

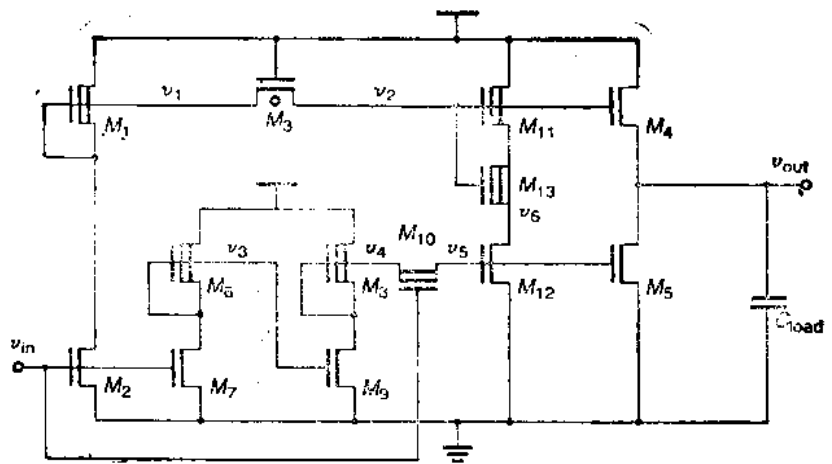


图5.19 第Ⅳ种情况：具有预驱动自举电容的自举超驱动器

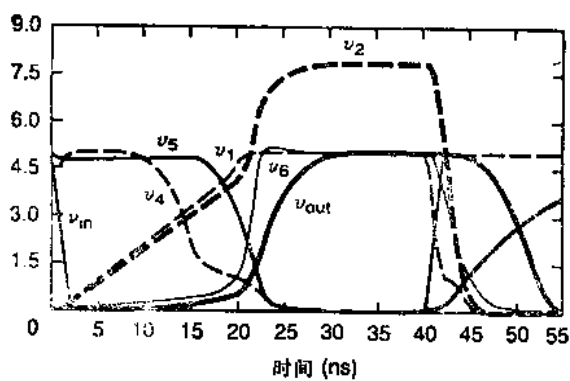


图5.20 第Ⅳ种情况的电压波形

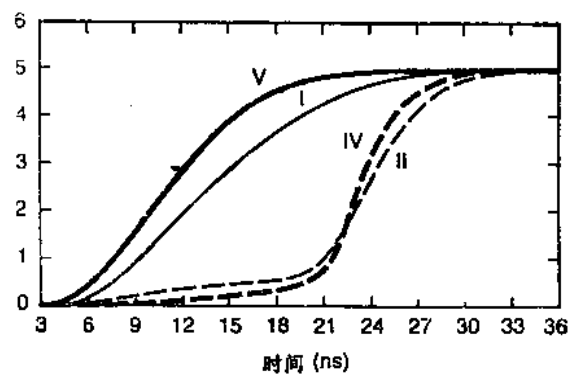


图5.21 不同电路型式输出波形的比较。请注意上升时间和输出首先开始变化时刻之间的折衷

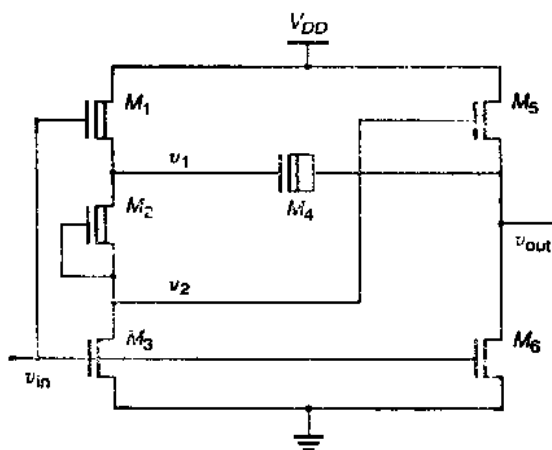


图5.22 第Ⅴ种情况：动态耗尽自举超驱动器

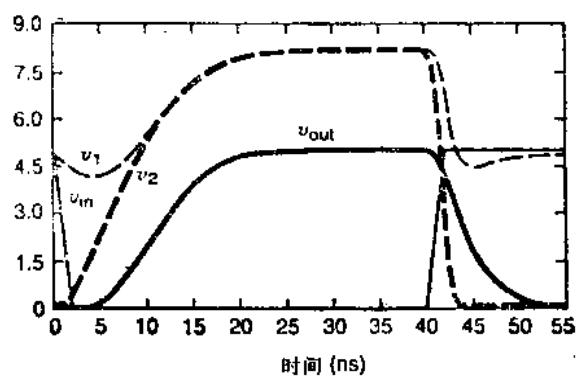


图5.23 第Ⅴ种情况的电压波形

中总免不了要有许多延时。在下边的技术中,自举电容总是已充好电备用。图 5.22 为第 V 种情况,它比前四种情况更加复杂[Knepper 78]。当输入为高电平时,除 M_6 外所有晶体管都导通。 M_1 至 M_5 的晶体管堆是有比的,晶体管 M_3 的栅较宽,因而使 v_1 为高 v_2 为低。这就意味着在电容 M_4 上的压降大约为 V_{DD} 。作为一级近似,其结果是该电容上的压降保持为常数。当输入电压下降时,晶体管 M_1 、 M_5 和 M_6 关断¹⁾。于是在电容 M_4 和驱动管 M_3 的栅极之间进行电荷分配,这个分配是通过“电阻” M_2 进行的。假如电容 M_4 是大的(典型情况下,它大于通常的自举电容),于是 V_{DD} 就加到了 M_3 的栅上。当输出升高时, v_2 和 v_1 处在同样电位,并升到 $2V_{DD}$ 。图 5.23 说明了这个电路的一些模拟波形。如情况 IV 中所述,可以用去掉负载与自举电容间耦合的办法来改进这个电路。

5.3 大扇入电路

当设计驱动大电容负载的电路时,为了使功耗最小或产生好的波形边沿,设计中要处理许多复杂问题。在工艺极限范围内,往往可以通过提高功耗得到更高的速度。式(5.28)表明,在负载电容增加时,若允许功耗按略大于线性规律增加,就可以做到总延时仅按对数律上升。高扇入电路则完全不同,PLA 多输入的或非阵列就是高扇入的示例。在高扇入情况下,逻辑门的负载电容(或固有电容)非常大,而充电或放电电流是有限的。我们有

$$dt = \frac{C}{I} dv \quad (5.43)$$

在 PLA 或非门的例子中,放电电流是由配置在矩阵中的下拉管的最大栅宽确定的。电容 C 决定于 PLA 中的字节数。参看式(5.43),假定 C 和 I 是固定的,减小 dt 的唯一办法是提高检测 dv 的灵敏度,这就意味着要采用不同类型的电路技术。解决这个问题包括两方面的任务。其一是设计一个电路来检测电压的微小变化,另一个任务是设计一个电路使负载保持在检测电路的高灵敏区。在本节,我们专门研究单端检测技术。在 5.5 节中,将介绍双端或差分检测技术。对于单端检测技术,我们要检测某些电压的绝对变化,而在双端检测技术中,是把信号电压与参考电压相比较,或者是把信号电压与反方向信号作比较。

一个检测小电压变化的方法是采用电荷分配放大器(共源-共栅连接),图 5.24 说明了这种结构,图 5.25 是典型的电荷分配放大器。我们假定 C_{big} 上的电压 v_σ 和 C_{small} 上的电压 v_β 都大于 V_{ref} 。 C_{big} 和 C_{small} 由 M_1 连通时的触发电压 $V_{\beta trigger}$ 为:

$$V_{\beta trigger} = V_{ref} - V_{Tf} \quad (5.44)$$

放大器的增益 a_0 由下式给出:

$$a_0 = \frac{V_{\sigma 0} - V_{\beta trigger}}{V_{\beta 0} - V_{\beta trigger}} \quad (5.45)$$

其中 $V_{\sigma 0}$ 和 $V_{\beta 0}$ 分别是 v_σ 和 v_β 的初始条件。我们曾经假定 $V_{\beta 0}$ 不受来自 C_{small} 电荷的干扰。可以把 $V_{\beta 0} - V_{\beta trigger}$ 看作是电路的高电平噪声容限。于是式(5.45)变成

$$a_0 = \frac{V_{\sigma 0} - V_{\beta trigger}}{NM_H} \quad (5.46)$$

1) 在这个电路中, M_3 是一个动态耗尽管,可以用一个增强型器件代替。

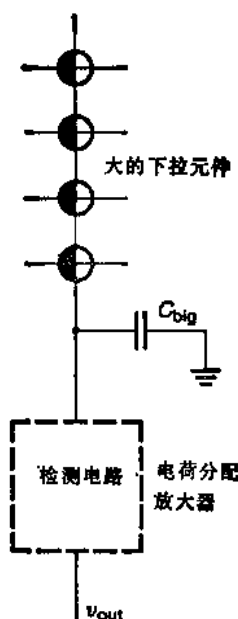


图5.24 带有电荷分配读出放大器的大输入单端电路

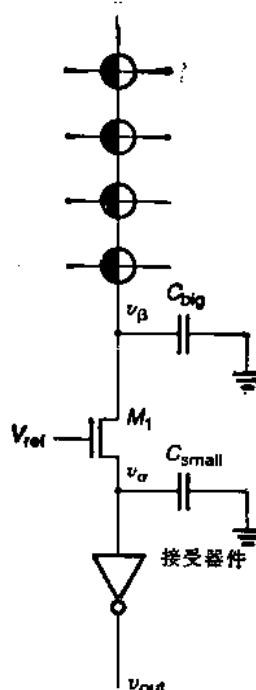


图5.25 电荷分配读出放大器

对于 5V 电源, NM_H 值的典型范围是 300mV 至 800mV。假定 V_{β_0} 小于 V_{DD} , 则最大增益 a_{0max} 变为

$$a_{0max} = \frac{V_{DD}}{NM_H} \quad (5.47)$$

从式 (5.43) 可知, 延时正比于噪声容限, 因而可以得出噪声容限和速度之间的主要折衷考虑。如果不考虑噪声容限, 则可以改进检测电路的速度。图 5.26 给出了 v_{α} 和 v_{β} 作为时间函数的典型曲线。

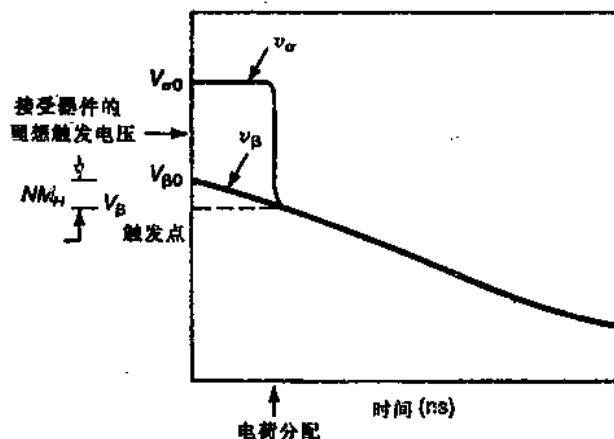


图5.26 电荷分配读出放大器的电压波形

图 5.27 是多种建立 V_{β_0} 和 V_{α_0} 方法中的一种。我们用这个电路作为示例是为了指出器件二级效应的重要性。请注意, 晶体管 M_1 和 M_2 上的全部电压是一样的。在这种情况下, 是利用短沟道效应得到 V_{β_0} 与 $V_{\beta_{trigger}}$ 之间的差别的。当然, 有许多更直接的方法预充 V_{α_0} 和 V_{β_0} , 但这里指出的方法能非常好地与工艺参数的变化进行统调。在设置 V_{ref} 值时存在折衷, 如果保持

V_{ref} 为低值, 则接收 v_{α} 的器件就可以简化, 而且此时增益 a_0 是高位。另一方面, 在 PLA 中, 当 V_{β_0} 高时 I_{ref} 为最大。

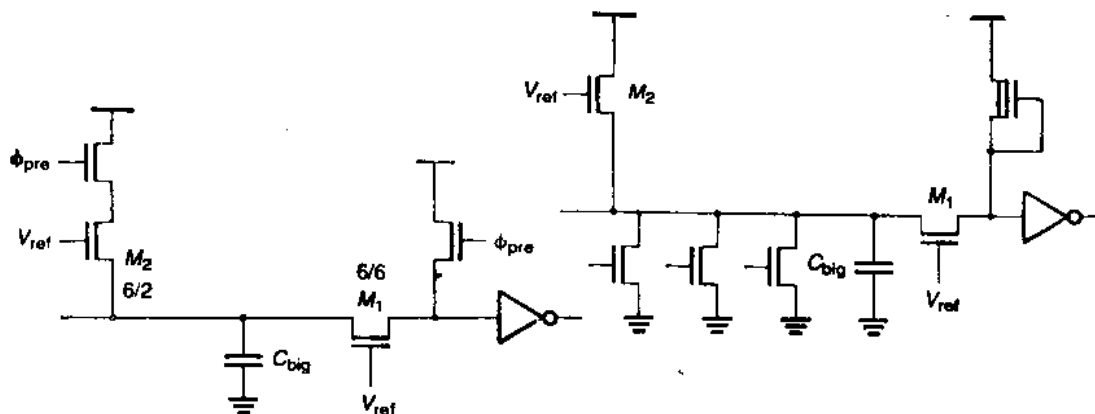


图5.27 利用短沟道效应实现的电路。该电路的噪声容限取决于短沟道效应引起的 6/2器件阈值电压的降低量

图5.28 用晶体管 M_2 限制 C_{big} 上电压摆幅的静态电路

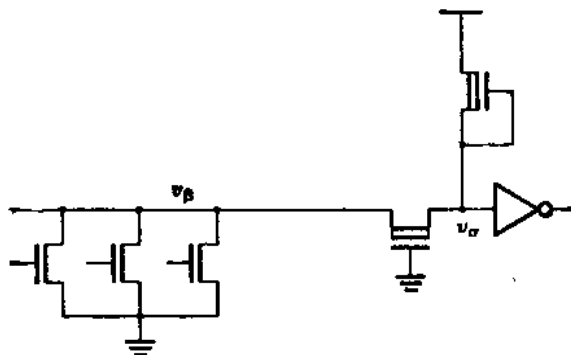


图5.29 第二种静态电路

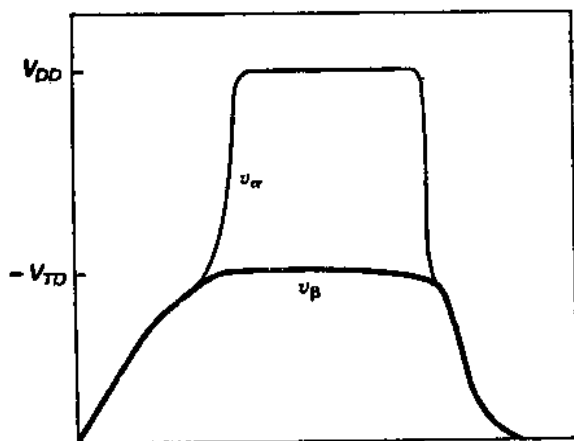


图5.30 在图5.29中所示电路的电压波形

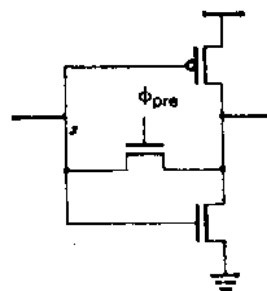


图5.31 对 V_{inv} 有补偿的电路

如果将或非门用于ROM，就是说如果在同一时刻只有一个下拉管导通，我们就可以设计 M_2 管将电压箝位到 v_{β} 。这样做可使位线不会被拉得太低。在不能用预充管把 v_{β} 回拉上去的静态电路中，这一点是很重要的，这样可以降低逐次读出之间的时间要求。保持小的电压摆幅可使容性耦合噪声减至最小。图5.28说明了这种设计。通过 M_2 的电流为 $(V_{DD}-v_{\beta})$ 平方的函数。如果在同一时刻只有一个下拉管导通，则 v_{β} 不会太低。

图 5.29 给出了这种电路的另一个有用的变型。图 5.30 是它的典型开关波形。

加快电路动作的另一种单端技术是使电路预充到 V_{is} 值。在图 5.31 中简单地说明了这一点。如果采用双轨逻辑 (dual rail logic)，这种技术可使电路动作特别快。在双轨逻辑中，一旦在某一时刻产生一个信号，则它的互补信号也会同时产生。在这些电路中，必须仔细地考虑噪声问题。

例 5.1 参考电压源的设计。

在 MOS 电路中，设计一个好的参考电压源，是非常困难的。通过对电源电压进行分压，往往可以获得足够好的参考电压源。图 5.32 给出了得到参考电压低于 V_{DD} 的两种电路。与图 5.32(b) 中的电路相比，图 5.32(a) 中的电路对 V_{TD} 的变化更不敏感。对于栅源相连的耗尽型负载器件，假定晶体管处在饱和区，则

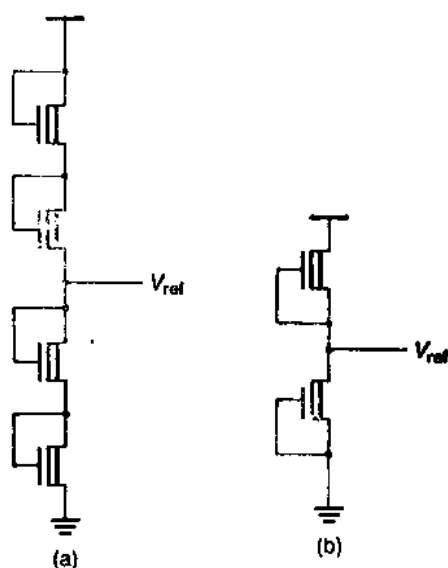


图 5.32 参考电压源。(a) 对 V_{TD} 变化不灵敏的电路；(b) 对 V_{TD} 变化敏感的电路

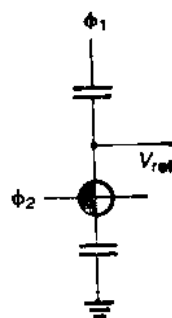


图 5.33 使用电容分压电路的参考电压源

$$I = K' S V_{TD}^2$$

对于同一器件，如果将栅漏相连，则

$$I = K' S (V_{out}^2 - 2V_{out}V_{TD})$$

其中 V_{out} 就是 V_{DS} 。让我们看一个例子，若 $V_{out} = 2V$ ， $-4V \leq V_{TD} \leq -3V$ ，在第一种情况下， I 变化 44%，而在第二种情况下， I 只变化 20%

在动态电路中，用上述第二类技术产生参考电压是很有用的。在这种电路中，可以采用如图 5.33 所示的电容分压电路。

5.4 双稳态

存在增益和正反馈的任何地方都会出现双稳态。在 VLSI 中，很多地方都使用双稳态电路，它包括有静态锁存器，触发器和施密特触发器¹⁾。

1) 在某些无谓之处，也会出现双稳态。例如，CMOS 电路中的自锁效应 (latch-up)。

静态锁存器

锁存器的作用是记忆。有两种基本类型的锁存器——静态锁存器和动态锁存器。在噪声不超过某固定幅度时，只要接通电源，一个静态锁存器就能记忆一位状态。这是因为静态锁存器包含有功率增益元件，它能连续地恢复代表状态的完整信号。为了对不太理想的电路元件和环境进行补偿，就必须具有功率增益。动态锁存器并不恢复代表状态的信号，因为在那里没有功率源。在某些技术中，可以制作能长时期存贮信息的存贮元件。例如，在超导体中能成年地存贮信息。在 MOS 中，动态存贮是通过电容上的电荷存贮实现的。对于典型的 MOS 数字电路工艺，动态存贮时间对 τ 的比值在 10^8 至 10^9 之间，这取决于温度和其它一些因素。有效逻辑电平附近的扰动不能恢复的情况下，动态锁存器与传输门逻辑是类似的。而静态锁存器则与恢复逻辑门相似。

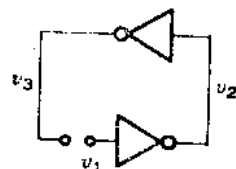


图 5.34 双稳态电路

最简单形式的静态锁存器包含有一对交叉耦合的反相器，如图 5.34 所示。双稳态锁存器有三个准静态解，如图 5.35 所示。在该图中，是用图解法求解的。由于对称性，存在一个两漏端电压为同一电压 V 的网络静态解。如果锁存器是双稳态的，在这个对称的工作点处，电路将是不稳定的。 M_1 和 M_2 （假定二者相同）都是增强型器件，都处在饱和状态，其栅电压与漏电压一致。因为通过 G_L 的电流等于通过晶体管的电流，所以有

$$(V_{DD} - V)G_L = \frac{K'S}{1 + \delta}(V - V_{TH0})^2 \quad (5.48)$$

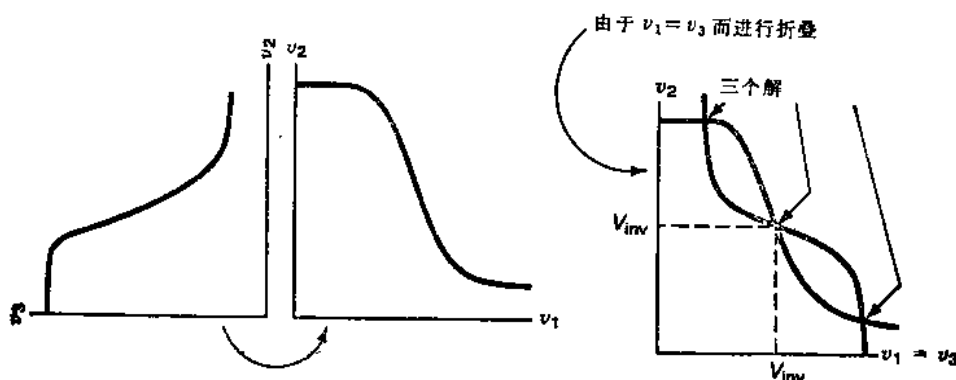


图 5.35 用图解法确定出三个静态解，其中只有两个点是稳定的

求解这个二次方程，我们得到

$$V - V_{TH0} = \frac{G_L(1 + \delta)}{2K'S} \left\{ \left[1 + 4(V_{DD} - V_{TH0}) \frac{K'S}{G_L(1 + \delta)} \right]^{1/2} - 1 \right\} \quad (5.49)$$

如果图 5.36 中电路的小信号回路增益大于 1，则该锁存器就是双稳的。回路是由两个同样偏置的反相器组成的，其中每个反相器的增益都必须大于 1，这是锁存器成为双稳所必须的条件。根据这一点，对于双稳态电路则有

$$g_m > g_L \quad (5.50)$$

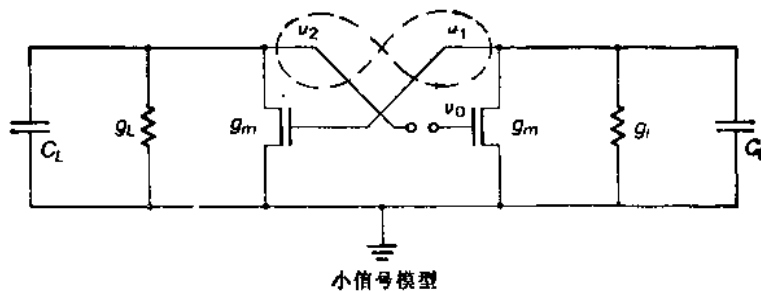


图5.36 无源负载交叉耦合锁存器的小信号模型

其中 g_L 是负载（可能是非线性的）的小信号导纳， g_m 为：

$$g_m = \frac{2K'S}{1+\delta}(V - V_{th0}) \quad (5.51)$$

合并式 (5.49)，(5.50) 和 (5.51) 则有

$$g_L < G_L \left\{ \left[1 + 4(V_{DD} - V_{th0}) \frac{K'S}{G_L(1+\delta)} \right]^{1/2} - 1 \right\} \quad (5.52)$$

简化后，可得双稳态的条件如下：

$$g_L \left(2 + \frac{g_L}{G_L} \right) < 4(V_{DD} - V_{th0}) \frac{K'S}{1+\delta} \quad (5.53)$$

对于电阻

$$1 = \frac{g_L}{G_L} \quad (5.54)$$

可以通过式 (5.49) 反映 V 和 G_L 间的联系。确定工作点后，同样可以把 g_L 和 G_L 联系起来。图 5.37 表明了 g_L 和 G_L 之间的关系。

已经有了标准的技术方法对交叉耦合锁存器的小信号和大信号动态特性进行分析。下边我们首先看看小信号分析。参看图 5.36，可有

$$g_m v_1 = - \left(g_L + C_L \frac{d}{dt} \right) v_2 \quad (5.55)$$

和

$$g_m v_2 = - \left(g_L + C_L \frac{d}{dt} \right) v_1 \quad (5.56)$$

我们要找出电压差值 $v_2 - v_1$ 的幅度。定义 Δv 如下：

$$\Delta v \equiv v_2 - v_1 \quad (5.57)$$

从式 (5.56) 减去式 (5.55)，得到

$$\frac{d}{dt} \Delta v = \left(\frac{g_m - g_L}{C_L} \right) \Delta v \quad (5.58)$$

或

$$\Delta v(t) = \Delta V \exp\left(\frac{g_m - g_L}{C_L} t\right) \quad (5.59)$$

其中 ΔV 是初始电压差。式 (5.59) 证实了我们以前的结论, 即如果 $g_m > g_L$, 则交叉耦合锁存器是双稳的。从式 (5.59) 可以看出, 如果 $g_m > g_L$, 则 Δv 随时间而增大。若 $g_m < g_L$, 则 Δv 随时间而减小。 $\Delta v(t)$ 达到某一特定值 ΔV_{switch} 的时间由下式给出:

$$t_{\text{switch}} = \frac{C_L}{g_m - g_L} \ln \frac{\Delta V_{\text{switch}}}{\Delta V} \quad (5.60)$$

从式 (5.60) 可见, 当 ΔV 趋于零时, t_{switch} 趋于无穷大。在一个有限的时间内, 锁存器状态的不确定性会引起同步问题。关于这一点, 我们将在第六章再进行讨论。对于快速开关, 要求 C_L 小而跨导 g_m 大。我们定义 $C_L / (g_m - g_L)$ 作为锁存器的特征时间常数 $\tau_{\text{switch}}^{1)}$ 。当 C_L 主要决定于晶体管电容和 $g_m \gg g_L$ 时, 可以找到 τ_{switch} 的最小值。对于这种情况, 可有

$$\tau_{\text{switch}} \geq \frac{2}{3} \frac{L^2}{\mu (V_{DD} - V_{TB})} \quad (5.61)$$

这个小信号分析, 提供了所需的有关静态锁存器的大多数信息。

为了在考虑了大信号效应的情况下讨论锁存器, 可以采用相平面分析技术 (phase plane analysis)。我们有

$$C_L \frac{dv_2}{dt} = I(v_1, v_2) \quad (5.62)$$

和

$$C_L \frac{dv_1}{dt} = I(v_2, v_1) \quad (5.63)$$

在这里

$$I(v_1, v_2) = I_{pu}(v_1, v_2) - I_{pd}(v_1 - v_2) \quad (5.64)$$

$$I_{pu}(v_1, v_2) = (V_{DD} - v_2) G_L \quad (5.65)$$

为了采用相平面分析技术, 我们取电压导数的比值。可有

$$\frac{dv_2}{dv_1} = \frac{I(v_1, v_2)}{I(v_2, v_1)} \quad (5.66)$$

式 (5.66) 右边的分子和分母分别表示当 $dv = 0$ 时两个反相器的静态传输特性。如图 5.38 所示, 在相空间中, 可以很容易地找到 $\frac{dv_2}{dv_1}$ 为零和 $\frac{dv_2}{dv_1}$ 为无穷大的点。由于对称性, 当 $v_1 = v_2$ 时, $\frac{dv_2}{dv_1} = 1$ 。

以时间作为独立参变量, 一些可能的 v_1 和 v_2 轨迹示于图 5.38 中。

当驱动大电容负载时, 比如大的寄存器堆作为负载或锁存器通过传输管输出的情况下, 有可能偶然地使锁存器翻转。对图 5.39 中六管 RAM 单元进行简单解释就可以看出这个问题。为了往单元中写 1, 字线 (连到两个传输管栅极的连线) 电位升高, 且同时

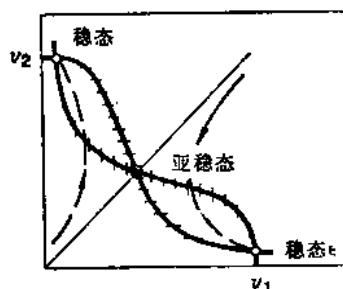


图 5.38 双稳态锁存器大信号特性的静态空间解

1) 在此计算中, 寄电容实际上是很重要的。

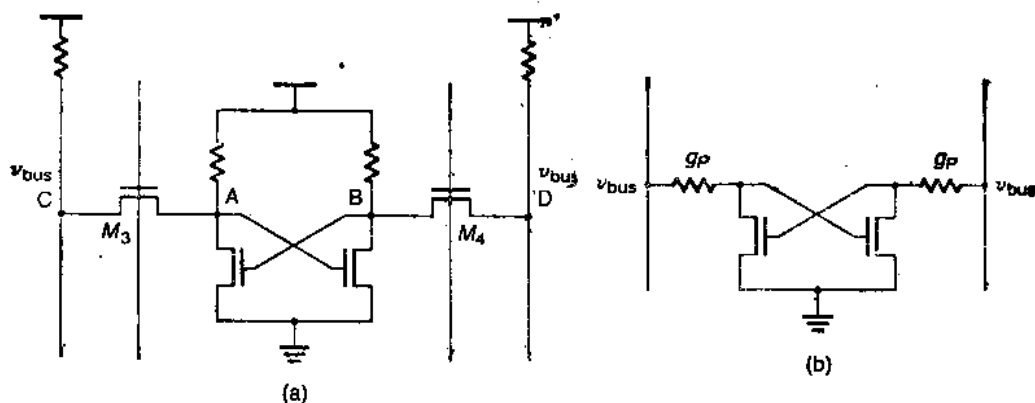


图5.39 静态MAR单元表示。(a) 存储晶体管；(b) 小信号模型

抬高左边位线电位而降低右边位线电位。当字线电位为低时，数据被锁存在单元中。为了实现读出，两条位线都被预充到高电位，而后使字线电位为高，则单元将迫使一条位线成为低电平。这时正是单元丢失存储信息的危险时刻。当传输管 M_3 和 M_4 导通时，节点 A 和 B 分别被连接到节点 C 和 D 上。在被寻址的上一位寄存器读“0”的地方，使节点 C 和 D 以比总线对晶体管上拉响应更快的速度下拉时，会出现“图案”效应 (patterning effect)。由于这种效应，使节点 C 和 D 的电压略低于 V_{DD} 。在出现这种情况时，锁存器不丢失原来的状态是十分重要的。对于传输管 M_3 和 M_4 大小的要求是有矛盾的。一方面我们希望把它们做得非常大，以便使节点 C 和 D 处大的总线电容能很快地被充电或放电，而另一方面，在晶体管 M_3 和 M_4 难以先导通的时候，等效负载和电源电压决定于传输的小信号电导 g_P 和总线上的电压 v_{bus} (假定 $g_P \gg g_L$)。在这种情况下，如果环路增益小于1，则会丢失所有的信息。式(5.53)可以用来确定用 g_P 和 G_P 代替 g_L 以及用 G_L 和 v_{bus} 代替 V_{DD} 的必要条件。另外一个重要的效应是当 M_3 和 M_4 先导通时，电荷耦合到单元里边。在实际电路设计中，在这种理论分析之后还要进行大量的电路模拟。可以用分析所得到的理解来指导这些模拟。

施密特触发器

常常用施密特触发器把变化缓慢的信号或变化不规则的信号转换为陡变信号。通常是在芯片的输入级见到施密特触发器。 n MOS和CMOS施密特触发器的原理图如图5.40所示。对于输入由低变高时CMOS和 n MOS施密特触发器触发点的分析是同样的。作为一级近似，可以认为 M_{N2} 开始导通时出现触发点。我们将分析 M_{N2} 的电流为零时的触发点。如果输入触发电压为 $V_{trigger}$ ，则反馈电压 V_{FN} 为

$$V_{FN} = V_{trigger} - V_{TN} \quad (5.67)$$

其中 V_{TN} 是 n 沟晶体管的阈值电压¹⁾。对于触发点处的输入电压，下拉管 M_{N1} 处在饱和区的边缘。 V_{FN} 决定于 M_{N3} 与 M_{N1} 之比。求解触发电压，可有

$$V_{trigger} = \frac{V_{DD} + \sqrt{\beta} V_{TN}}{1 + \sqrt{\beta}} \quad (5.68)$$

1) 由于这是粗略的分析，所以忽略了体效应。把体效应包括进去并不难。

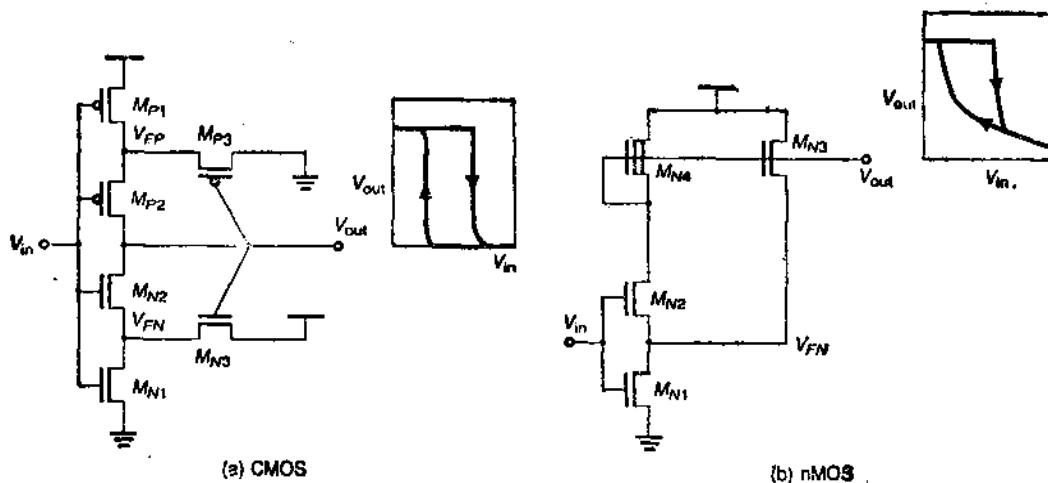


图5.40 施密特触发器。(a) 用CMOS实现的电路及其静态传输曲线；(b) 用nMOS实现的电路及其静态传输曲线

其中

$$\beta = \frac{S_{TN1}}{S_{TN3}} \quad (5.69)$$

显然, $V_{TN} < V_{trigger} < V_{DD}$ 。从式(5.68)可以看出一个要点, 即触发电压对工艺参数是不灵敏的。在低 β 值时, 对 V_{TN} 的要求也不太严格, 而且触发电压也和耗尽型器件(或P沟器件)的阈值电压无关, 甚至该阈值电压并不出现在这些一级近似的公式之中。

超过触发电压之后, 沿反馈回路的增益就大于1了。这将导致电路突然地转到一个新的稳定状态。由于对称性, 在CMOS施密特触发器中, 对于由高变低的输入信号, 输出也是陡变的。然而要注意, 对于由高变低的输入信号, 电路的触发电压将决定于P沟器件而不再决定于n沟器件。

施密特触发器的明显迟滞特性如图5.40中的直流传输函数所示。施密特触发器可以用作电平转换。然而, 在nMOS施密特触发器中, 必须留心它对由高变低的输入信号呈现出的软开关响应特性。在开关动作开始之前, nMOS施密特触发器的输出电压可升高并超过 V_{OL} 值。施密特触发器也可以与锁存器和电阻分压器做在一起。应当注意, 由于增益-带宽的折衷, 通常施密特触发器的速度要比同样功耗反相器的速度慢一些。因为几个宽尺寸器件在比较长的时间内同时导通, 所以在微功耗CMOS电路中要考虑施密特触发器中的抢电流问题。图5.41中给出了一个nMOS施密特触发器的版图。

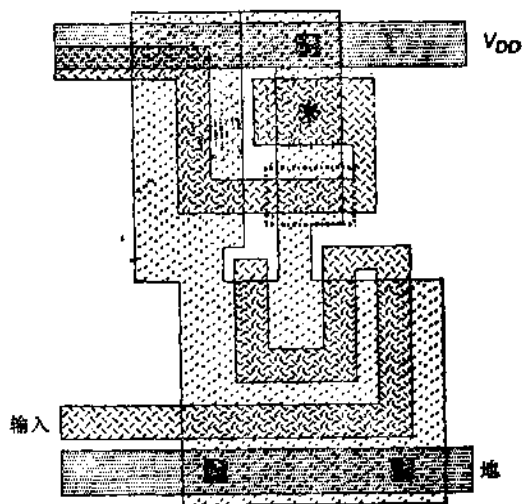


图5.41 施密特触发器的版图

例5.2 定时信号发生器。

由于施密特触发器具有如此稳定的触发电压，所以可以用它产生一个相当稳定的延时。如果把触发电压设定得高一些，则触发电压仅随 V_{DD} 明显变化。如果用一个接电源

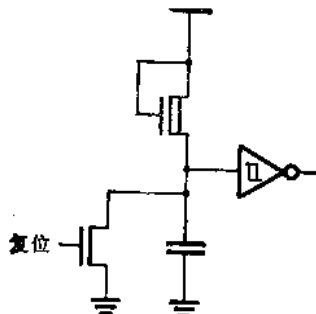


图5.42 延时发生器。迁移率的变化是造成不精确的主要原因

V_{DD} 的电阻给电容充电，并且在电容上电压超过某阈值时用施密特触发器进行检测，则由于延迟时间的限定，抵消了 V_{DD} 的变化，而只剩下电容和电阻的变化了。在一个已定的工艺线上，栅电容可以严格控制。为了获得良好的电阻，可以用外接电阻或片内的晶体管。在nMOS电路中，把耗尽型器件的栅接到 V_{DD} 上，就可以成为一个合适的电阻。采用这种结构，通过把器件尺寸画大的办法，可以使电阻变化与迁移率变化达到同样量级。这类电路示于图5.42中。请注意，在这里抵消了栅电容的变化。在CMOS电路中，可以用易控的在线性区工作的器件限定充电电流。如果不再需要这种“不正常”的器件，就可以使用电流镜。

用易控的在线性区工作的器件限定充电电流。如果不再需要这种“不正常”的器件，就可以使用电流镜。

5.5 读出放大器

由于读出放大器在存贮技术中的重要作用，恐怕没有对其它MOS电路进行过比它更为细致的研究和优化了。读出放大器用作检出并放大的输入电压差值。我们将研究三种最重要的读出放大器。

静态差分放大器

在模拟运算放大器中，用静态差分放大器作为基本前置级。图5.43说明了它的基本电路结构。其中电流源常由处在饱和状态的长沟道晶体管构成。它由电流镜提供电流偏置以便于和加工工艺变化进行统调。当输入 v_1 和 v_2 相等时，由于对称性，流过 M_1 和 M_2 的电流都是 I 。

由于输入电压的不平衡使二管之一抢了电流。在极端情况下，一个晶体管被关断，使其漏电压为 V_{DD} 。与此同时，流过另一个晶体管的电流为 $2I$ ，它的漏电压为 $V_{DD} - 2I/G_L$ 。为了使放大器的增益最大，希望尽可能使 M_1 和 M_2 处在饱和区的工作范围大些。当它们处在饱和区时，可以写出流过输入晶体管的电流如下：

$$I_1 = K' S (v_1 - v_x - V_T)^2 \quad (5.70)$$

和

$$I_2 = K' S (v_2 - v_x - V_T)^2 \quad (5.71)$$

求解 $v_2 - v_1$ ，我们得到

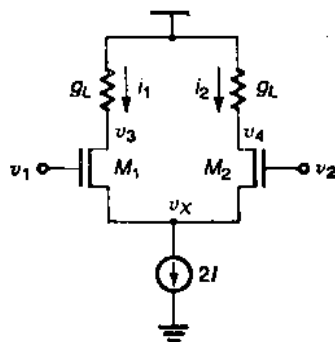


图5.43 静态差分放大器的电路型式

$$v_2 - v_1 = \sqrt{\frac{I_2}{K'S}} - \sqrt{\frac{I_1}{K'S}} \quad (5.72)$$

从式 (5.72) 可以看出, 输入动态范围可由下式给出,

$$V_2 - V_1 = \pm \sqrt{\frac{2I}{K'S}} \quad (5.73)$$

当输入电压相等时, 如果 M_1 和 M_2 处在饱和状态, 这些晶体管的跨导则为

$$g_m = 2\sqrt{K'SI} \quad (5.74)$$

为了找出从输入到输出的增益, 我们将电路线性化, 并进而研究它对两输入电压差的响应以及它对平均输入电压的响应。我们定义 Δv_{in} 为:

$$\Delta v_{in} \equiv v_2 - v_1 \quad (5.75)$$

定义 Δv_{out} 为

$$\Delta v_{out} \equiv v_4 - v_3 \quad (5.76)$$

进入到电流源的小信号电流和必为零。即

$$0 = i_1 + i_2 \quad (5.77)$$

其中

$$i_1 = g_m(v_1 - v_x) \quad (5.78)$$

$$i_2 = g_m(v_2 - v_x) \quad (5.79)$$

我们还有

$$g_L v_3 = -i_1 \quad (5.80)$$

$$g_L v_4 = -i_2 \quad (5.81)$$

合并式 (5.78) 和 (5.80), 合并式 (5.79) 和 (5.81), 得到

$$g_L v_3 = -g_m(v_1 - v_x) \quad (5.82)$$

$$g_L v_4 = -g_m(v_2 - v_x) \quad (5.83)$$

从式 (5.83) 减去式 (5.82) 得到

$$\Delta v_{out} = -\frac{g_m}{g_L} \Delta v_{in} \quad (5.84)$$

可以看出, 只要维持 M_1 和 M_2 处于饱和态, 由于 g_m 是电流镜决定的, 所以电路增益和输入电压的准确值无关。在大信号情况下, 显然增益不再是常数。增益较大时的最大输入电压范围可由式 (5.73) 给出。把式 (5.82) 和 (5.83) 相加, 我们可以得到

$$(v_3 + v_4) = -\frac{g_m}{g_L}(v_1 + v_2 - 2v_x) \quad (5.85)$$

根据式 (5.77), (5.78) 和 (5.79), 可有

$$0 = v_1 + v_2 - 2v_x \quad (5.86)$$

合并式 (5.85) 和 (5.86), 得到

$$0 = v_3 + v_4 \quad (5.87)$$

上式告诉我们, 电路的共模增益为零。电流源使平均输出电压保持为常数。

由于晶体管 M_1 和 M_2 存在体效应, 也由于晶体管的非理想性, 使得电路的共模增益

不等于零。这种电路的一个更重要的限制是很难保持其中所有晶体管工作在饱和区。这一点限制了输出电压的摆幅。

时钟差分放大器

像静态差分放大器一样，时钟差分放大器也用了—个电源 [Stein72]，但在这种电路中，电流源是随时间变化的。图 5.44 给出了时钟差分放大器。在这个电路中没有上拉器件。电容上电压间的差值被交叉耦合晶体管放大。由于输入电压不相等，所以其中一个晶体管流过更多的电流。由于正反馈，使具有较低电压的输入节点将被拉到更低的电位。我们的目标是设计一个电路，使较高输入电压时基本上没有变动，而在输入电压较低时，将节点强拉至地电位。

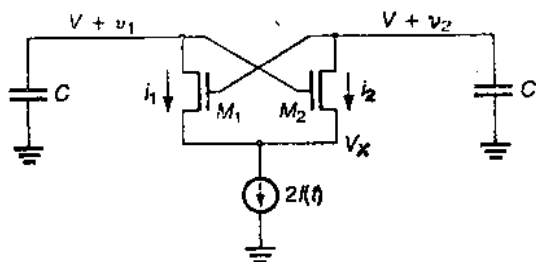


图 5.44 时钟差分放大器的电路型式

我们首先研究差分电压为零的情况。这就是网络对共模的响应。可有

$$I(t) = -C \frac{dV(t)}{dt} \quad (5.88)$$

和

$$I(t) = K'S[V(t) - V_{TB} - V_x(t)]^2 \quad (5.89)$$

求解 $V(t)$ 可有

$$V(t) = V_0 - \frac{1}{C} \int I(t) dt \quad (5.90)$$

对于 $V_x(t)$ ，有

$$V_x(t) = V_0 - V_{TB} - \sqrt{\frac{1}{K'S}} - \frac{1}{C} \int Idt \quad (5.91)$$

当 $V - V_{TB} \geq V_x$ 时，这些公式是适用的， V 和 V_x 均随 $I(t)$ 的上升而降低。电流越大 $V(t)$ ，下降越快。

为了弄清该电路是怎样实现读出放大功能的，我们分析一下差分扰动叠加到大信号共模响应上的电路响应。假定扰动电压 $v_1 = -\Delta v/2$ 和 $v_2 = +\Delta v/2$ 。小信号电流之和必为零。可有

$$i_1 + i_2 = 0 \quad (5.92)$$

其中

$$i_1 = -\frac{C}{2} \frac{d}{dt} \Delta v \quad (5.93)$$

$$i_2 = +\frac{C}{2} \frac{d}{dt} \Delta v \quad (5.94)$$

我们还有

$$i_1 = -g_m \left(\frac{\Delta v}{2} + v_x \right) \quad (5.95)$$

$$i_2 = +g_m \left(\frac{\Delta v}{2} - v_x \right) \quad (5.96)$$

合并式 (5.93) 一式 (5.96), 得到

$$C \frac{d}{dt} \Delta v = g_m \Delta v \quad (5.97)$$

与静态差分放大器不同, 此处 g_m 是时间的函数。可以把式 (5.97) 重写为

$$\frac{d}{dt} \ln \Delta v = \frac{g_m}{C} = \frac{2}{C} \sqrt{IK'S} \quad (5.98)$$

或

$$\Delta v = \Delta v_0 \exp \left[\frac{2}{C} \sqrt{IK'S} \int \sqrt{I(t)} dt \right] \quad (5.99)$$

当然, 式 (5.99) 适用于小信号 Δv 。在此电路中, 最终一个晶体管将关断, 而另一个晶体管将脱离饱和态。

作为电路设计者, 他面临的问题之一是选择适当的电流波形 $I(t)$ 。我们知道, 通常 $V(t)$ 是下降的而 $\Delta v(t)$ 则是上升的。我们希望凡是电容上具有较高初始电压的, 其电压保持不变, 而电容上具有较低初始电压的, 则电压被驱动到地电位。就共模和差模信号而言, 这种转换使 $\Delta v(t)$ 的增大比 $V(t)$ 的下降快两倍。这样一来, $v_2(t)$ 将保持恒定, 而 $v_1(t)$ 将以 $d\Delta v/dt$ 的速度下降, 其平均值将按此速率的一半下降。从数学角度上看, 我们要求

$$\frac{\frac{d}{dt} \Delta v}{\frac{d}{dt} V} \approx -2 \quad (5.100)$$

从式 (5.88), 可有

$$\frac{d}{dt} V = -\frac{I}{C} \quad (5.101)$$

而从式 (5.97) 或 (5.74), 则有

$$\frac{d}{dt} \Delta v = \frac{2\Delta v}{C} \sqrt{IK'S} \quad (5.102)$$

合并式 (5.100) 至式 (5.102), 得到

$$I(t) = K'S(\Delta v)^2 \quad (5.103)$$

由于 $|\Delta v|$ 是随时间而增大的 (我们希望如此), 所以式 (5.103) 意味着在初始时 I 将缓慢上升而随后将快速上升。如果 $I(t)$ 停止增加, 则 Δv 就不可能增大, 并且 $V(t)$ 将在其差值被放大之前减弱。此时 Δv 增大的速度 [参看式 (5.98)] 将与电流强度有关。因此, 希望在读出放大器中两边的电流都减到零的限制下, 使 $I(t)$ 尽可能大些。可以近似地用图 5.45 中的电路或用类似的电路去说明理想的电流波形。如果所有的读出放大器都具有大体相同的平均输入电压, 则不论是读“1”还是读“0”, 对于所有放大器 v_x 都应为相同。既然所有 v_x 节点都处在同一电位, 所以可以把它们连在一起而获得良好结果。这样一来, 时钟下拉管就可以被所有的放大器公用, 并且还可以把它们从单元中移出来。然而这样有可能把电流 I 弄得太大。电流取的太小会降低电路的工作速度, 但并不破坏电路功能。为了使速度最佳, Lynch 和 Boll [Lynch 74] 曾经指出, 可以把处在高电位的连线也部份地降到地电平, 同时电流波形的一般形状仍保持不变, 这样做可

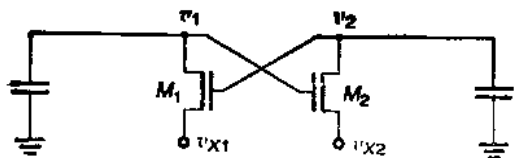


图5.47 另一种差分放大器

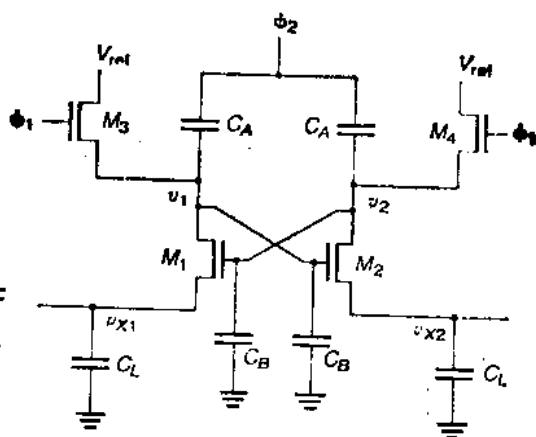


图5.48 差分电荷前级放大器

在 ϕ_1 降低后 ϕ_2 升高。像5.5节的时钟差分读出放大器一样，为了使灵敏度最高， ϕ_2 必须缓慢地升高。

可以完全仿照时钟读出放大器的分析进行这个电路的分析。下边我们只做小信号分析。使电容的充电电流相等，可有

$$-v_{x2}C_L = v_2C \quad (5.104)$$

和

$$-v_{x1}C_L = v_1C \quad (5.105)$$

其中 C 是漏节点的总电容。它被定义为

$$C \equiv C_A + C_B \quad (5.106)$$

我们还有

$$C \frac{dv_1}{dt} = -g_m(v_2 - v_{x1}) \quad (5.107)$$

$$C \frac{dv_2}{dt} = -g_m(v_1 - v_{x2}) \quad (5.108)$$

合并上边两个公式，得

$$\tau_A \frac{d\Delta v}{dt} = \Delta v \quad (5.109)$$

其中

$$\Delta v = v_2 - v_1 \quad (5.110)$$

$$\tau_A = \frac{C}{g_m \left(1 - \frac{C}{C_L} \right)} \quad (5.111)$$

注意，在 $C_L \gg C$ 时，式(5.109)趋近于式(5.97)。从 $v_{x2} - v_{x1}$ 到 $v_2 - v_1$ 的初始电压增益为 $-C_L/C$ 。

如果输入电压 v_{x1} 和 v_{x2} 十分接近，则最终差分电压的最大值为 V_{T8} 量级。由于这个原因，往往把这种电压看作是预放大级。通常把它和时钟读出放大器组合到一块，以便提供较大的最终电压差。其工作时序为：

ϕ_1 高 预充 v_{x1} 和 v_{x2} ，

终端读出放大器的原理图和时序图如图 5.49 所示。

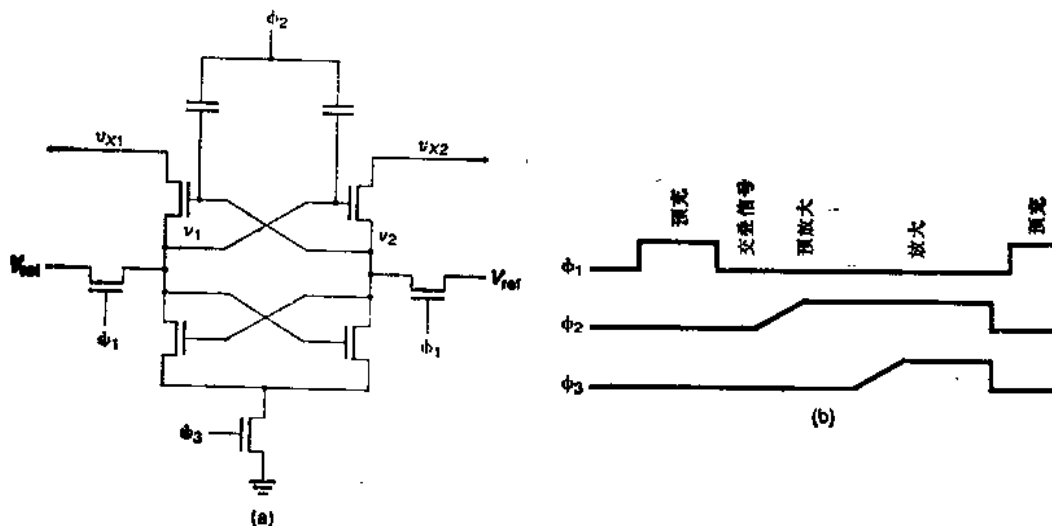


图5.49 (a) 将差分电荷电路和时钟读出放大器组合在一起的电路; (b) 该电路的波形图

图 5.50 是一个单端 CMOS 电荷分配读出放大器^[1]，它采用了 5.3 节的电路型式。作为一个实例，曾用 CMOS 工艺制成了这块电路。这个电路可以用在大容量 ROM 的输出，噪声容限设计为 0.5V。在该电路的版图设计时，必须小心地将参考电压的噪声耦合至最小。图 5.51 是该电路电压波形的模拟结果。从图可以看出，p 沟预充管的关断

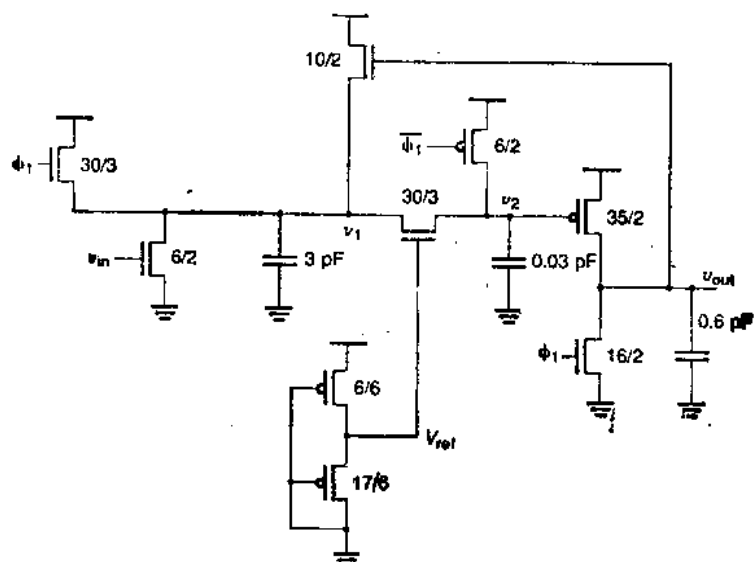


图5.50 帶有快速恢復反饋晶体管的CMOS电荷分配读出放大器

1) 该电路是由C. Bamji设计的。

使预充节点电位超过 V_{DD} 十分之几伏。这类预充电路被普遍采用，为了实用，还必须进行工艺设计，以便在各种条件下不会产生锁定效应。

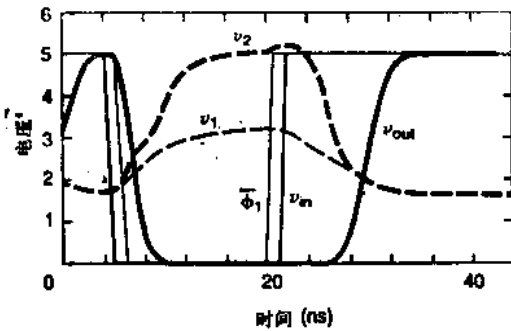


图5.51 图5.50所示电路的电压波形

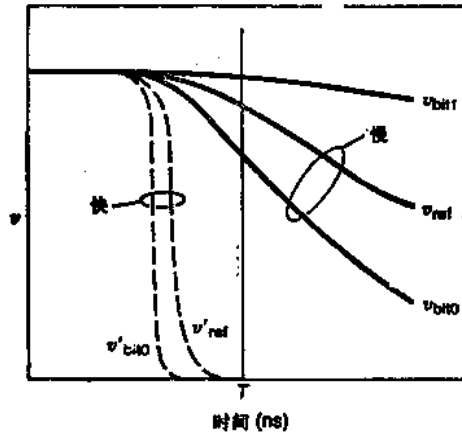


图5.52 具有虚单元线的读出放大器波形。使一个电路在快速和慢速两种情况下都能工作是很困难的

有许多其它方法可构成ROM的读出放大器，其中包括采用 5.5 节的差分读出放大器。

有两种基本的产生参考电压的方法。其中之一是可以用一个固定的参考电压源，如图 5.50 中所用的参考电压源，这是一个好办法。另一个方法是在ROM中加入一个附加列线（位线）。这种方法的原理是使这个附加列线满载，这样一来，其电压下降速度就会比其它列线电压下降慢一些。该电路的波形如图 5.52 所示。图中同时给出了慢速情况和快速情况。如果我们分析慢速（实线点）波形，就可以看出，不论选中的位线是读“1”（ v_{bit1} ）还是读“0”（ v_{bit0} ），在时刻 T 都能很好地响应读出。这种技术的限制在于快速芯片，在时刻 T 快速芯片读出操作的性能可能是很差的，这是因为 v'_{bit0} 和 v'_{ref} 都将减小到地电位。

肯定有办法使这个预想设计实现，然而会有点困难（也许是十分困难）。在这里有一个经验可以吸取。在这个应用中，如果采用一个固定的参考电压，就只需要集中于慢速工作。与此相反，如果采用了图 5.52 的动态参考电压，我们就可以在整个工作范围内实现设计目标。一般讲，采用动态参考电压并非坏事，难办的是必须同时解决两个复杂的而又互相有关的问题。与解决两个无关的问题相比，这往往是极端困难的。对于全局设计方法，可以把某个问题分割成几个小问题，即在一个时刻只关心一件事情，然后加以解决。这就是所谓的分而治之。

5.6 电平转换

在选定了将在电路主要部分使用的设计方法和有效电压电平之后，会遇到一些例外的情况。不论是芯片内部和外部，有许多地方的电平与电路主要部分的电平不兼容，所以就需要移动有效电平。与此同时，往往也要求放大信号。最普遍的情况就是把TTL信

号电平转换到5V MOS逻辑电平。

对TTL电路规定的有效输入电平为 $V_{IL}=0.8V$ ，和 $V_{IH}=2V$ 。对于MOS电路来讲，上述的低电平太高而高电平太低。让我们首先研究一下高电平。在耗尽型负载的nMOS电路中，2V的 V_{IH} 将要求电路具有很大的 β 比值。就其本身而言，这并不成为问题。唯一棘手的乃是由于 V_{IH} 如此之低，从而使最坏条件的 V_{TE0} 可能是高了而不是低了。这一点已在第4.4节中讨论过了。

TTL电平转换的第二个问题是 V_{IL} 值会超过增强型晶体管的阈值电压。这就意味着从高 β 比反相器输出的高电平将达不到 V_{DD} 值。这就要求另有一个第二电路来达到规定的 V_{IL} 低电平值。这样一来，简单的TTL兼容输入级要包含两级反相器。第一级高 β 比反相器用来调整 V_{IH} 高电平值，第二级反相器用来调整 V_{IL} 低电平值。如果标准的工艺文件中不包含特定的TTL最坏条件文件，则需要用一个噪声最坏条件的文件来校正最坏条件的 V_{TE0} 值。

另一个典型的问题是使信号降低到非常靠近 V_{DD} 值。例如，在5.3节中，电荷分配放大器的输出集中在相当高的电平处。在这种情况下，源跟随器就很有用了。图5.53中示出了典型的源跟随器。只要晶体管处在饱和状态，它的电流就正比于 $(V_{GS}-V_{TE})^2$ 。为了使漏到源的电流保持不变， $V_{GS}-V_{TE}$ 必须是常数。在这个工作区中，可有

$$v_{out} = v_{in} - V_{TE0} - \sqrt{\frac{(1+\delta)I}{K'S}} \quad (5.112)$$

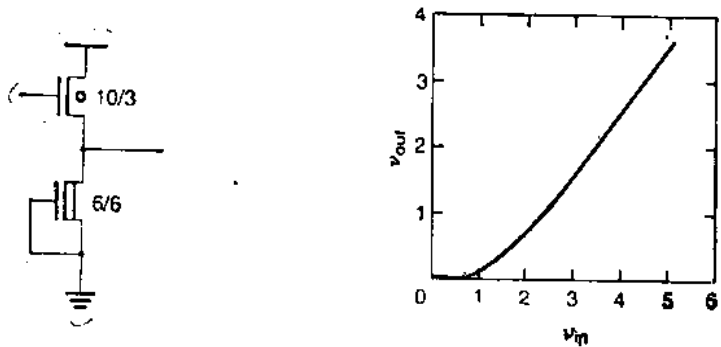


图5.53 源跟随器的静态传输特性

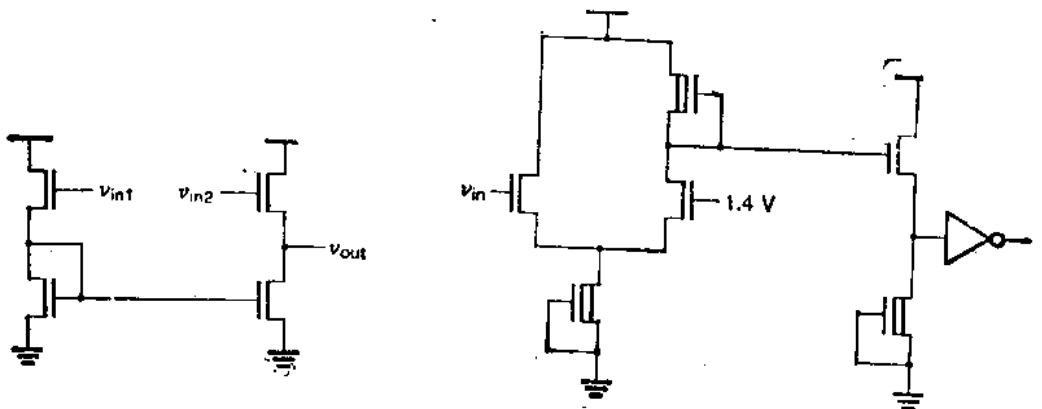


图5.54 差分信号转为单端的转换器

图5.55 TTL输入缓冲器

图 5.53 还说明了简单的源跟随器的传输函数。可以用电流镜来偏置下边的晶体管。电流镜中的晶体管应该采用长沟器件，以便使工艺偏差的影响减至最小。图 5.54 说明了设计 MOS 模拟电路常用的一种电路技术，用这个电路可以实现双端到单端的转换。在高频时，源跟随器的动态特性是很好的，这是因为 C_{gs} 的作用降低了输入电容并且向前馈送了信号。

读出放大器（参看 5.5 节）还能用作差分信号的电平移动，或者能把差分信号和参考电压作比较。图 5.55 说明了采用读出放大器的简单 TTL 输入缓冲器。

5.7 输入保护

VLSI 工艺是从芯片内部电压的角度优化的，该电压等于或小于 5V。在与芯片的接口处，会遇到电压幅度几乎高三个数量级的情况。所以必须设计输入保护电路（常被称为“避雷器”）以保护芯片。这类电路的设计和分析是很困难的，它经历了一连串的改变。

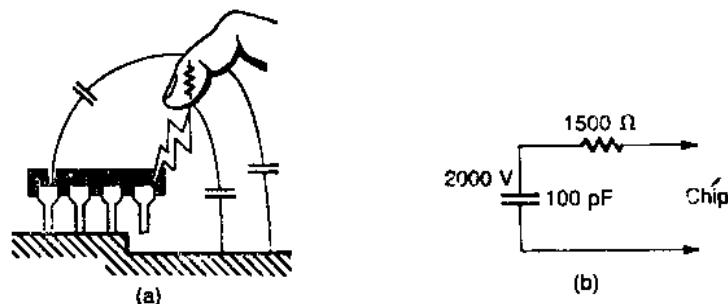


图 5.56 人的静电模型。（a）物理结构；（b）等效电路

图 5.56 说明了人的简单静电模型。它包括一个与 $1.5\text{k}\Omega$ 电阻相串联的 100pF 电容，该电容被充电到 2000V 。这个模型还有许多变型。例如，一些厂家所试验的输入保护电路，用了一个被充到 300V 的电容而没有用串联电阻。我们来看一下手指接近 MOS 芯片的情况。当达到某一距离时，充了电的手指将与芯片产生弧光放电。电压低的时候，比如说 300V ，手指需要离芯片非常近才发生放电，这时电容也要高一些，比如说 300pF 。当电压高的时候，手指离芯片较远就会发生弧光放电，这时即使电容较低些也会在手指的较大面积范围内发生弧光放电，这将影响放电电阻。我们的目的是要设计一个电路，能够允许人接触封装好的芯片引出端，而不毁坏这个元件。

箝位和放电电荷的滤除

在 2000V 时，存储在电容中的能量 $(\frac{1}{2}CV^2)$ 是很大的，约为 0.2mJ 。静电放电可以通过好几种方式破坏 MOS 器件，我们仅仅讨论其中的两种。最明显的机构是因施加过量的电场而导致的栅氧化层击穿。通常二氧化硅的极限电场大约为 $7 \times 10^8 \text{V/cm}$ 。假定电场完全均匀， $300\text{\AA}$ 的栅氧化层能承受的最高电压大约为 21V 。第二种破

坏机构是 p-n 结或多晶硅连线的热烧毁。硅的热容量约为 $0.7\text{J}/(\text{g} \cdot \text{K})$ 。为了对在人身电容中究竟存贮了多大的能量得到一个感性认识，我们计算一下它可熔化的硅体积。硅的熔点是 1415°C ，如果开始时硅处在 15°C ，所以需考虑 1400°C 的温度升高。由上述可知，芯片外部具有的能量为 $0.2 \times 10^{-3}\text{J}$ 。假定这个能量的一半耗散在硅中，可以算出，一次静电放电可以熔化 $45000\mu\text{m}^3$ 体积的硅¹⁾。

最通用的输入保护结构，包含有一个 n^+ 有源区长条，它的终端连到一个增强型晶体管的漏区，该晶体管的源区和栅连接到地端。这个输入 n^+ 线的典型串联电阻为 500 至 1000Ω ，该晶体管的栅宽在 80 至 $100\mu\text{m}$ 之间。图 5.57 示出的这种结构已被证明是合理的和有效的。

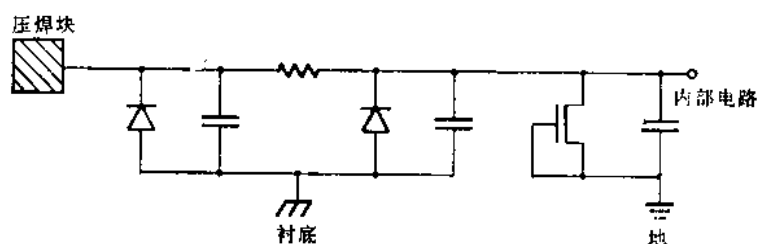


图 5.57 nMOS 输入保护器件的电路模型

在此， n^+ 输入电阻用一个 Π 形结构来模拟，它实际上是一个 RC 和二极管制成的分布长线。这种结构有两个作用。其一，对于输入尖峰脉冲，它可作为低通滤波器；其二，它可对输入信号进行箝位，使它的负值不小于衬底电位与一个二极管正向压降之和。该低通滤波器必须设计成不与输入信号所必需的带宽相抵触。换句话说，它的截止频率越低越好。其中的增强型晶体管，也有好几个作用。首先，它对内部电压进行箝位，使其负值不小于地电位加上一个阈电压值。在芯片具有负 V_{DD} 衬偏时，正是这个晶体管而不是 p-n 结二极管首先提供箝位。这个晶体管的第二个作用是限制内部电压的正向范围。利用它可以产生穿通和雪崩击穿两种限压机构。

问题是穿通电压强烈地依赖于晶体管的有效沟道长度 L 。结深也会通过二级效应影响 V_{PT} 值。对于短沟道器件，为了实现保护，它们必须具有容许的低穿通电压值，在大多数工艺线上这个穿通电压势必很难控制。

在正常条件下，为了有效地进行输入保护（在 40V 量级），p-n 结二极管的雪崩击穿电压有些过高。可以通过栅接地的晶体管降低这个电压。这个栅极可以使 p-n 结中的反偏电场集中，使得在较低的电压下发生雪崩击穿，这个效应是重复的，并且在 15 至 17V 左右就可引起雪崩击穿。为了使雪崩击穿超过穿通而成为主导的效应，晶体管的栅必须长一些。这种技术的唯一缺点是雪崩击穿发生在靠近器件的表面区，因而保护晶体管的栅要宽一些，以便耗散所要求的能量而不引起烧毁。

当输入保护电路失效时，靠近输入端的 p-n 结二极管部分往往要被烧毁。由于衬底是电阻很低的材料，所以在采用外延层的 CMOS 工艺中，这个问题特别严重，因为在那里对最大电流很少限制。一些设计者曾采用多晶硅作输入保护电阻，这种结构不再有 n^+ 结构中的 p-n 结烧毁问题。另一方面，多晶硅布置在衬底的上面使它处在热导很

1) 这仅仅是一个粗略的计算，在此假定了比热与温度无关，还忽略了热传输。尽管如此，它还是能够对这个问题的概念。

差的状态,这将引起十分严重的问题。多晶硅必须非常宽以便增加它的整个耐热能力。这样作了之后,该结构也被证明是有效的。然而,在这方面,还没有 n^+ 电阻技术那样多的经验。

金属栅场区晶体管是另外一种保护结构(这种晶体管是经常出现的,通常被看作是寄生器件)。图 5.58 示出这种器件。在有大的正输入电压瞬变过程的情况下,它必然工作在线性区。我们有

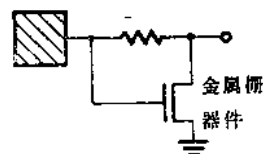


图 5.58 用厚氧化层场区金属栅晶体管作为输入保护器件

$$v_{\text{internal}} = 1 + 2K'_p S R (\overline{v_{\text{spike}} - V_{TF}}) \quad (5.113)$$

为了使之有效,形状因子 S 必须为 100 或更大些。 K'_p 是关于场区晶体管的,其增益直受边缘场的严重影响。这类场区晶体管并不是自对准的。在这类结构中,CVD 氧化层的击穿可能会成为问题。

在输出端或时钟输入端不用串联电阻,这是因为在这些引线端无法容忍由这个电阻导致的性能退化。对于时钟端,常规的输入增强型晶体管栅宽必须超过 $200\mu\text{m}$,以提供附加的保护。在输出端,往往由输出晶体管提供保护。最小的总沟道宽度的典型值为 $500\mu\text{m}$ 。

对 CMOS 的附加考虑

对于 CMOS 电路,如果不采用适当的隔离环,输入保护电路就可成为自锁的诱发器。例如,在 5.7 节所讨论的保护电路中,用一个栅和源接地的 n 沟道增强型晶体管作为箝位器件。再次提醒一下,本书讨论的是 n 阱 CMOS 工艺。

对一个正向变化的放电过程,晶体管进入雪崩模式,由于雪崩过程产生电子空穴对,一部分放电电流将由这些载流子形成。在源和漏附近,耗尽区中的电场迫使绝大多数电子趋向漏端或源端,在那里它们将被收集并从电源线或信号节点流出。

多子空穴将流过衬底,且通过连到衬底的接触端流到衬底偏置电源。对于 CMOS 电路(n 阱),衬底偏置往往处在地电位。这个多子电流将使保护电路附近的衬底电位高过地电位。

若 n^+ 源区或漏区附近处在地电位,而衬底电位超过约 400mV ,则源与衬底的 $P-n$ 结被足够地正偏,从而引起前向导通并导致少数载流子(电子)向衬底注入。附近的包含有 P 沟器件的阱区就可以收集这些电子,假如在这个阱中的 pnp 双极型三极管被导通,则由初始 n^+ 源区(它是横向 nnp 双极型晶体管的一部分)所形成的回路就可被激励而发生自锁。

输入端对保护电路的负向放电,可以被两种机构箝位。首先考虑 MOS 二极管箝位模式。如同在 $nMOS$ 的示例中那样,当漏端被拉向负值,它的作用就如同源跟随器中晶体管的源区。这就使漏区(现在是源区)被箝位在比地电位低一个阈值的电位上。现在再来考虑 $P-n$ 结箝位模式。在 CMOS 中,衬底往往是接地的,而不像 $nMOS$ 电路,其衬底被偏置在负电位。因此,相对于衬底而言, n^+ 漏区的结将变为正向偏置。在此,电流要比第一种机构大一些。因而 $P-n$ 结箝位模式往往是主要的。

第二种箝位方法的问题在于总的箝位电流包含有注入到衬底的少数载流子。其扩散

长度（对于外延工艺而言，或叫做有效横向扩散长度）在 $100\mu\text{m}$ 量级，因而距离保护电路几百 μm 的范围内，都会受到这些少数载流子的干扰。

如果在相当大的范围内存在着潜在的能自锁的 pnpn 回路，而且这些载流子被该回路中 pnp 晶体管的阱区所收集，则该回路就会被驱动到自锁状态（pnp 垂直阱晶体管将首先导通，随之它将驱动相邻横向 npn 晶体管导通）。

考虑一下这个例子，我们就可看到正的和负的放电（任一个都是可能的）都将引起衬底干扰，即使是距离几百微米，这个干扰也可以触发灵敏的自锁回路。在放电时流过的电流可以很大，1500 V 通过人身电容和约 1500Ω 电阻放电，将产生 1 A 的峰值电流。

考虑一个正向放电，在那里，电流的一部分是雪崩产生的多子电流，它通过衬底流到 V_{ss} 或地的接触端。假定总放电电流的 $1/3$ 表现为雪崩产生的多子衬底电流。典型的垂直衬底电阻为 200Ω （对于外延衬底），这个箝位电流流过该电阻（在正的雪崩情况下），将导致衬底电位升高 60 V。当然，不会升高那么多，这是因为其它 n 阱和 n^+ 源区接到 V_{DD} 上，所以将衬底箝位到 V_{DD} 之后，就会引起结的正偏。这些前向偏置的结构将产生少数载流子，从而增加衬底的电导。不幸的是这些少数载流子也能诱发自锁效应。无论如何，这种干扰是强烈的，且可扩展到几百微米的范围。

这个问题的解决办法是减弱 ESD 产生的自锁效应的干扰。常常用两种保护环减弱在瞬变过程中衬底内可能存在的大干扰电流（可参考 2.3 节中讨论的自锁效应）。一种方法是使少数载流子衰减，用一个把保护单元整个包围起来的阱区长条收集从负偏 n^+ 发射区注入的少数载流子。如 2.3 节中所述，这种类型的中间收集区是非常有效的，在采用外延工艺时，可使少子电流的幅度衰减 4 至 6 个数量级 [Troutman 83]。如果不是外延工艺，这种阱区结构只能使电流幅度衰减 1 至 2 个数量级。环到阱区可以做得宽一些，以便更有效地阻止没有被收集的从阱下深处过来的少数载流子（由于硅片厚度约为 $500\mu\text{m}$ ，而阱的深度只有 $3-5\mu\text{m}$ ），所以载流子可以容易地从阱下边通过。

另外一个技术方法则是衰减多数载流子。采用外延工艺时，外延层下边是高掺杂衬底，从而使干扰随距离迅速衰减。对于非外延工艺，借助于环绕单元的接触条，将靠近单元的衬底表面连到地（或 V_{ss} ）。这个表面接触将阻止表面电位升到使附近电路进入自锁的状态。

从这些概述中可以得出有关 CMOS 保护电路的一般性结论。如果芯片在工作的系统环境中发生静电放电，虽然保护电路可以有效地吸收放电的能量，周围的一部分电路仍可进入自锁状态。如果电路没有破坏，则只有通过先切断芯片电源，随后再加上电源的办法去掉自锁状态。在这种情况下，为了防止自锁现象，必须加入隔离环，以便减弱通过衬底到芯片其它形成可自锁 pnpn 回路处的干扰。

因为存在互补 p 沟器件，在保护电路设计中 CMOS 比 nMOS 电路更加灵活。有时使用的另一种保护电路示于图 5.59 中，多晶硅电阻吸收了大部分放电能量，采用了 p-n 结箝位， p^+ 扩散区形成的“二极管”与 n 型阱区实际上成为一个垂直 pnp 晶体管，它以阱区作为基区，以衬底作为集电区。只要阱收集环加在 n^+ 区和衬底二极管的周围，并收集由于放电而注入到衬底的少数载流子，则它与阱区就成为横向 npn 晶体管。 n^+ 扩散区作为 npn 的发射区，衬底成为基区，阱收集环成为 npn 集电区。

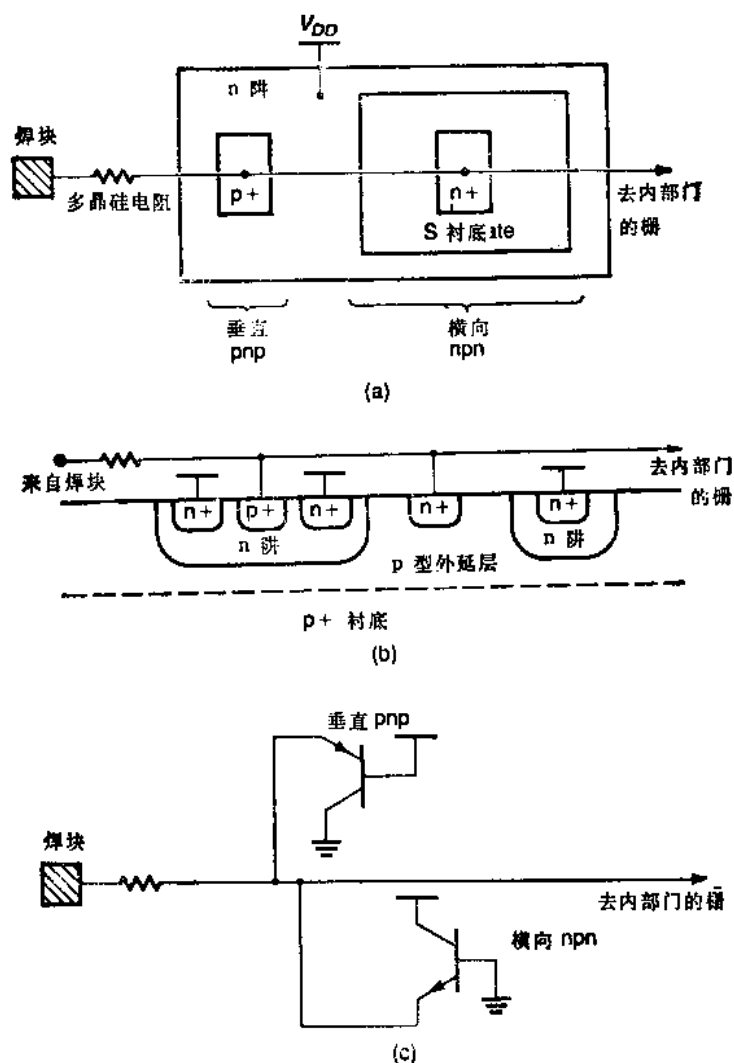


图5.59 采用 p⁺和 n⁺扩散区的CMOS ESD保护电路。通过p-n结使之箝位到 V_{DD} 和地电位。两个结成为垂直的 (pnp) 和横向的 (npn) 双极型晶体管。这与nMOS中所见为简单的二极管不同。(a) 版图俯视图, (b) 剖面图, (c) 等效电路

5.8 衬偏电压的产生

除片外电源以外, 芯片内部产生的电源给我们增加了一个重要的自由度, 它常被用于 MOS 电路中, 可以产生高于 V_{DD} 或低于地电位的电压。在 nMOS 芯片中最常用的是 V_{ss} 负电压发生器[Cheng 80, Huffman 77, Lee 79, Martion 80, Pashley 76, Puri 78]。其它应用包括 V_{GG} 发生器和为显示而用的高压驱动电路。自举是电压变换的一个简单例子。

将衬底接 $-V_{ss}$ 比接地有下述优点: 它可降低阈值电压对体效应的灵敏度, 提高穿通电压, 不需要减少衬底掺杂就能降低扩散区与衬底的结电容, 能够增加典型反相器的增益。它还可以降低由时钟控制的耗尽型晶体管的次开启漏电流, 并且防止芯片由于在输

入端有负脉冲电压尖峰而使衬底转入正偏, 这种情况在采用 TTL 外围电路时是常见的。如果提供了反馈, V_{BB} 电压还能补偿某些器件参数的变化。

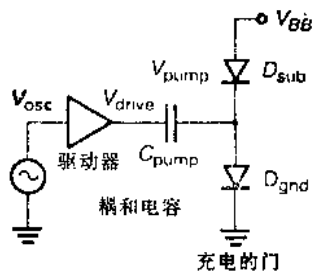


图 5.60 一种 V_{BB} 发生器的理想模型

用“泵”把电子从地节点抽取到衬底中, 可以产生 V_{BB} 负电压。由于衬底是一个大电容 (在 1000pF 的量级), 这就降低了它的电压。图 5.60 给出了 V_{BB} 发生器的一种典型结构。其中驱动器放大了由振荡器产生的 ac 信号, 并给电荷泵以电能, 该电能是通过电容 C_{pump} 耦合的。两个二极管 D_{gnd} 和 D_{sub} 控制电荷从衬底流出并流入到地节点。当电压 V_{pump} 接近它的峰值时, D_{gnd} 被正偏, 电荷被“泵”打入到地节点。此时另一个二极管 D_{sub} 是关断的。在周期的负半周, V_{pump} 接近它的最负值 D_{gnd} 被关断, 同时 D_{sub} 把电荷从衬底抽取出来。当然, 电荷是不能流过电容器的。

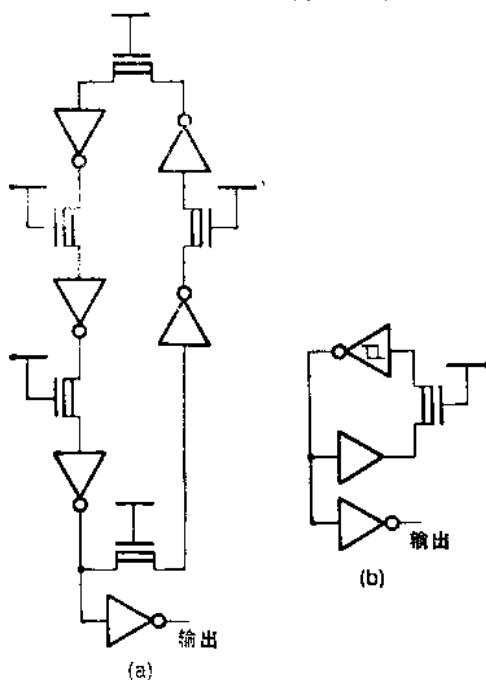


图 5.61 (a) 环振电路; (b) 施密特触发振荡器

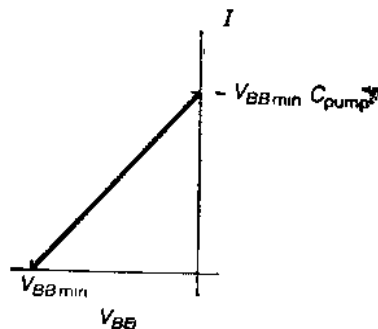


图 5.62 V_{BB} 发生器的理想 $I-V$ 特性

振荡器

衬偏发生器用的振荡器有两种电路形式。对于一个 N 为奇数的环形振荡器, 振荡频率 f 近似为 $(2N\tau_{delay})^{-1}$, 其中 τ_{delay} 是每一级的延迟时间。由于加在所有门上的负载是不等的, 所以上述关系只是近似的。 N 应该大于或等于 5, 否则电压摆幅就太小了。

施密特触发器连接成带有 RC 滤波器的回路, 可以用来代替环形振荡器。环形振荡器和施密特触发器电路示于图 5.61 中。典型振荡频率范围为 5 至 10MHz。由于实际原因, 往往把该振荡器的振荡频率锁定到系统的时钟上。

电荷泵二极管

V_{BB} 的理论最小值决定于 V_{pump} 的峰-峰值和两个二极管中的电压降。在周期的正

半周, 为了抽取电荷, V_{pump} 必须比地高一个二极管压降 (二极管 D_{gnd})。在周期的负半周, V_{pump} 必须比 V_{ss} 低一个二极管压降 (二极管 D_{sub})。假定 V_{pump} 的最大峰-峰值电压小于 V_{DD} , 则 V_{ssmin} 大于或等于 $-V_{\text{DD}}$ 加两个二极管压降。由于漏电流和寄生电容的影响, 很难达到这个理想值。

由于 nMOS 工艺带来的限制, D_{gnd} 通常是用一个栅和漏接在一起的增强型晶体管做成的。 D_{sub} 可用这种方式实现, 或者用制造工艺技术自然提供的 p-n 结二极管实现。如果用了晶体管式的“二极管”¹⁾, 衬底就必须连到外部, 通常用一个芯片内部的焊点把 V_{ss} 连到衬底背面的接点上。这样做时, 必须小心保证芯片背面和内部焊点之间的良好接触。要注意, 在该处往往会存在 p-n 结二极管。

如果是理想的二极管, 则从衬底抽出的电流将是

$$I_{\text{avg}} = \Delta V C_{\text{pump}} f \quad (5.114)$$

其中 ΔV 是衬底电压 V_{ss} 与 V_{ssmin} (约为 $2V_{\text{T}} - V_{\text{DD}}$) 理论最佳值间的差值。当 ΔV 较大时, 即电荷泵首先启动时, 电流是大的。电荷泵必须能在所有 V_{ss} 电压 (在零和它的最小值之间) 下工作²⁾。 I_{avg} 随泵的频率 f 而增加。在图 5.62 中说明了它的理想响应。

当 $\Delta V C_{\text{pump}}$ 所表示的全部电荷确实是从衬底抽出时, 式 (5.114) 是有效的。对于理想二极管, 这一点往往是真实的, 但对于实际器件, 正向偏置器件的有限电阻大大地限制了电荷泵的效率。重要的问题是二极管的 RC 时间常数和泵电容值是大于 $1/f$ 还是小于 $1/f$ 。如果时间常数是短的 (对于理想二极管), 式 (5.114) 就是适用的而 $v_{\text{pump}}(t)$ 的波形为限幅脉冲波。如果时间常数是长的 (在 V_{ss} 接近 V_{ssmin} 时就是这种情况), $v_{\text{pump}}(t)$ 就近似为正弦波。

现在, 我们来研究上述第二种情况。集中注意 D_{gnd} 。可以看到, 只有 $v_{\text{pump}} > V_{\text{TE0}}$ 时电流才流通。在周期波的顶部附近, 可以用在最大值附近的泰勒级数很好地近似 $v_{\text{pump}}(t)$:

$$v_{\text{pump}} \approx V_0 - \frac{V_0}{2} \frac{\theta^2}{2} \quad (5.115)$$

其中 V_0 是 v_{pump} 的峰-峰值, 而 $\theta = 2\pi ft$ 。我们定义 $v_{\text{pump}} - V_{\text{TE0}}$ 的最大值为 ΔV 。在 V_{ss} 达到它的理论最小值时 ΔV 为零。在这种情况下, v_{pump} 超前 V_{TE0} 的相角在 $\pm 2\sqrt{\Delta V/V_0}$

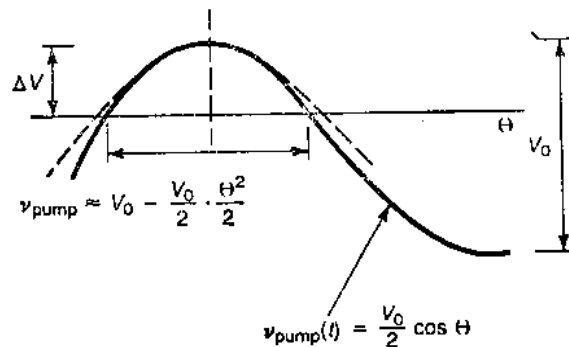


图 5.63 电压波形。用抛物线来近似正弦波 $v_{\text{pump}}(t)$ 的顶部

之间, 如图 5.63 所示。通过 D_{gnd} 的电流由下式给出:

- 1) 因为 V_{ss} 往往是零, 所以该晶体管的阈值电压是低的。
- 2) 在某些 nMOS 工艺中, $V_{\text{ss}} = 0$ 时, 在低于 V_{DD} 的情况下, 短沟道器件实际上是穿通的。

$$i(t) = K' S \left(\Delta V - \frac{V_0 \theta^2}{4} \right)^2 \quad \text{对于} -2\sqrt{\frac{\Delta V}{V_0}} \leq \theta \leq 2\sqrt{\frac{\Delta V}{V_0}} \quad (5.116)$$

对上式进行积分, 以便找出一个周期中所抽出的电荷。我们有

$$Q = \int i(t) dt = \frac{2K' S}{2\pi f} \int_0^{2\sqrt{\Delta V/V_0}} \left(-\frac{1}{4} V_0 \theta^2 \right)^2 d\theta \quad (5.117)$$

或

$$Q = \frac{2}{5\pi f} K' S V_0^2 \left(\frac{\Delta V}{V_0} \right)^{5/2} \quad (5.118)$$

如果 v_{pump} 是一个方波而不是正弦波, 则有

$$Q' = \frac{K' S V_0^2}{2f} \left(\frac{\Delta V}{V_0} \right)^2 \quad (5.119)$$

从式 (5.118) 得到每一个周期抽出的平均电流为

$$I_{\text{avg}} = Qf = \frac{2}{5\pi} K' S V_0^2 \left(\frac{\Delta V}{V_0} \right)^{5/2} \quad (5.120)$$

可比较式 (5.120) 与式 (5.114)。衬底电压的平均变化率可由 $I_{\text{avg}}/C_{\text{sub}}$ 给出。这里的 C_{sub} 为衬底电容。可以把它看成负载电阻, 以粗略地模拟 I_{avg} 对 v_{pump} 的影响。我们有

$$R_{\text{load}} \approx \frac{\Delta V}{I_{\text{avg}}} \quad (5.121)$$

这个分析与导出式 (5.114) 的主要不同在于式 (5.114) 中假定波形被限幅了, 而在这个分析中没有这个假定。事实上, 在这个分析中, 我们假定了 $v_{\text{pump}}(t)$ 与 $i(t)$ 无关。这说明了两种情况的不同频率依赖关系。式 (5.114) 和式 (5.120) 说明了从电荷泵流出最大电流的限制。图 5.64 示出电荷泵的一般性能。

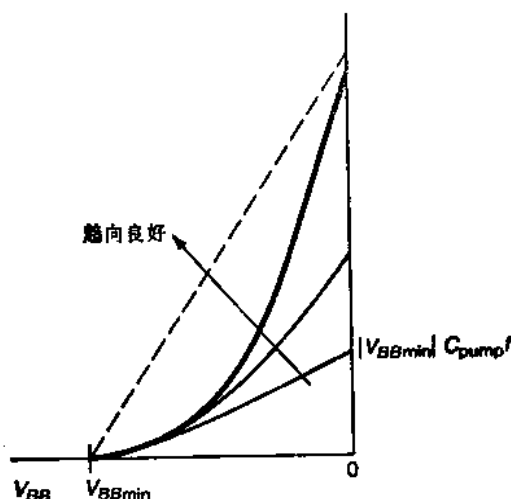


图 5.64 非理想 V_{BB} 发生器的特性

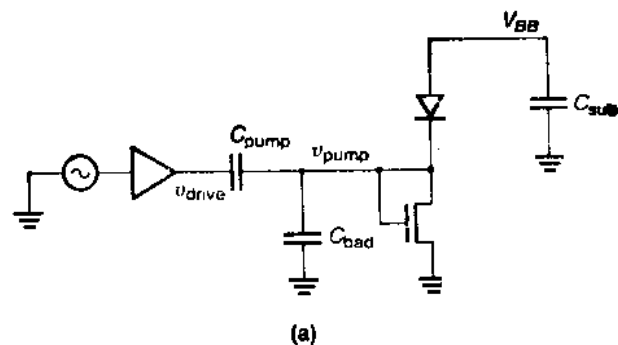
耦合电容器

耦合电容器 C_{pump} 把能量从驱动电路转换到电荷泵中去, 电荷泵把电子从较高电位送到较低电位处。图 5.65 给出了 V_{BB} 发生器, 其中重要的电容都有明显标记。 C_{bad} 是一个寄生电容, 其作用是降低 V_0 , 因而也降低了 V_{BB} 的最小值。当 $V_{BB} \approx V_{BB\text{min}}$ 时, 因为 $C_{\text{sub}} \gg C_{\text{bad}}$ 和 $2\pi f C_{\text{pump}} \gg G_{\text{load}}$, 所以

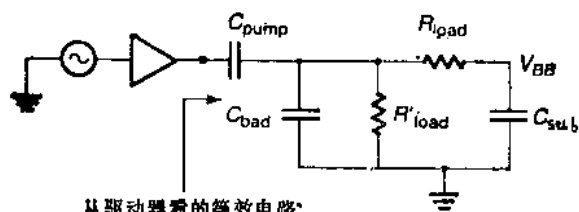
$$v_{\text{pump}} \approx \frac{C_{\text{pump}}}{C_{\text{pump}} + C_{\text{bad}}} v_{\text{drive}} + V_{\text{offset}} \quad (5.122)$$

如果 v_{drive} 的峰-峰电压为 V_{DD} , 则有

$$V_0 \approx \frac{C_{\text{pump}}}{C_{\text{pump}} + C_{\text{bad}}} V_{DD} \quad (5.123)$$



(a)



(b)

图5.65 V_{BB} 发生器的模型。(a)重要的电容和二极电路；(b)小信号模型

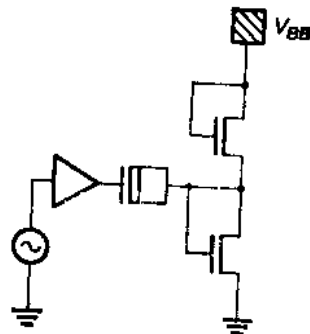


图5.66 用MOS晶体管作二极管和电容

显然，我们希望 C_{bad} 最小以便减小 $V_{BB,min}$ 值。为了使 V_{BB} 大些，要求 $C_{pump} \gg G_{load}/2\pi f$ 。而为了使 C_{bad} 减小，要把 C_{pump} 的版图画成源漏连在一起的大方块形耗尽型晶体管。通常要把晶体管的栅连接到电荷泵上。当 $v_{drive} = V_{th}$ 和 $v_{pump} = V_{TBO}$ 时，由于耗尽型晶体管处在低于阈值电压的状态，所以这一点往往是不可能的。如果出现这种情况，则沟道消失且 C_{pump} 的幅度下降几个数量级。因而常常把 C_{pump} 作成如图 5.66 中所示的形式。在此，晶体管的栅被连接到驱动器上。还必须注意，不要把 C_{pump} 晶体管的栅作的太长。因为这个晶体管的时间常数正比于该晶体管栅长的平方，所以晶体管的几何形状很容易作成对时钟的响应过慢。除非给予极大注意，否则在电路模拟中很难发现这个问题。 C_{pump} 的典型尺寸为 $200\mu m \times 200\mu m$ 。如果 $2\mu m$ 栅长晶体管的时间常数 τ 为 $0.2ns$ ，则 $200\mu m$ 栅长晶体管的时间常数 τ 就是 $2\mu s$ 。在晶体管内部加入一些接触孔就可以解决这个问题。二极管形状因子 S 的设计也存在折衷。当 S 增加时，式 (5.118) 中的电流上升，但是由于寄生电容 C_{bad} 的增加而使 V_{BB} 下降。

驱动器

仔细观察图 5.65 后可以看出，驱动器并不需要去驱动 C_{pump} 的等效负载，这是幸运的。当 $\Delta V \approx 0$ 时，驱动器应在有限范围变化。在这些条件下， R_{load} 非常大，而驱动器的等效负载是与 C_{bad} (或近似 C_{bad}) 相串联的 C_{pump} 。当电荷泵刚起动时， $\Delta V \approx V_{DD}/3$ ，而最小微分电阻近似为 $(2K'S\Delta V)^{-1}$ 。这个电阻是与 C_{bad} 并联，而与 C_{pump} 串联的。因为在起动时这个电阻是低的，所以初始负载近似为 C_{pump} 。这一次自然界对我们还是很好的，在大摆幅无关紧要时，在一开始见到的是大负载。随后，当大摆幅重要时，则看到小负载。

在稳态, 由衬偏发生器抽取的平均电流 I_{ss} 等于流入到衬底的漏电流。这个漏电流主要来自饱和负载器件注入到衬底的少数载流子 (热电子效应)。典型电流值为 $5\mu A$ 。

因为衬偏电压影响芯片的好几个电参数, 比如功耗和阈值电压, 所以可以通过衬偏发生器反过来控制这些参数。有些反作用本来就有。例如, V_{ss} 电压影响着少数载流子向衬底的注入, 而反过来这个少数载流子向衬底的注入又影响着 V_{ss} 电压。在所有反馈的系统中, 稳定性都可能成为问题。温度也影响 V_{ss} 电压。在低温下, “二极管”性能更好, 漏电流更小, 因而 V_{ss} 电压一般更负一些。通常 V_{ss} 对温度的正常依赖关系对阈值电压控制有不良影响, 但趋向于稳定耗尽型器件中的功耗。要注意, 当热电子效应为主要时, 在低温下衬底电流可能最高。

电荷泵版图设计技术

V_{ss} 发生器的性能依赖于许多版图设计技术。例如, 我们已经讨论过, 应当使 C_{nsd}/C_{pump} 最小。因为一个小的电位降都可能带来问题, 所以衬偏发生器的良好接地是很必要的。衬偏发生器往往安放在靠近地端焊块处, 并要用金属搭接扩散区和多晶硅区以使电压降最小。

是选取双极型 $p-n$ 结二极管作 D_{ss} 还是用 MOS 晶体管和焊点作 D_{ss} , 这需从多方面考虑。在商品化芯片中, 两种类型都采用过。 $p-n$ 结二极管并不需要外界焊块。在靠近有源电路处, 有向衬底注入电子的潜在问题。在该电路依赖于动态电荷存贮时, 这个电荷可能被消除掉。连接到 V_{DD} 的扩散保护环可能会有一些帮助。但是尺寸比是个不利因素。衬底约为 $250\mu m$ 厚, 而环区的深度大约只有 $1\mu m$ 。除这些电子, 它们想必会使衬底电压降低。另一个潜在问题是 V_{pump} 节点电荷的泄漏。由场氧化层上的多晶硅或金属线构成的厚场区晶体管是主要的漏电器件, 这是因为它们实际上是短沟道晶体管的

缘故。由于电路的电荷泵部分实际上在零 V_{ss} 值下工作, 使得这个问题更加严重。在芯片中这些部分场区晶体管的阈值电压是比较低的 (场区晶体管的体效应系数可能是 $50V^{1/2}$)。如果用晶体管实现 D_{ss} , 就可以用与整个区域相隔离的硅栅接地厚场区晶体管把电荷泵与电路的其它部分隔离开来。多晶硅是连到衬底的。有时用一个外接电容来稳定 V_{ss} 电压以抗噪声。

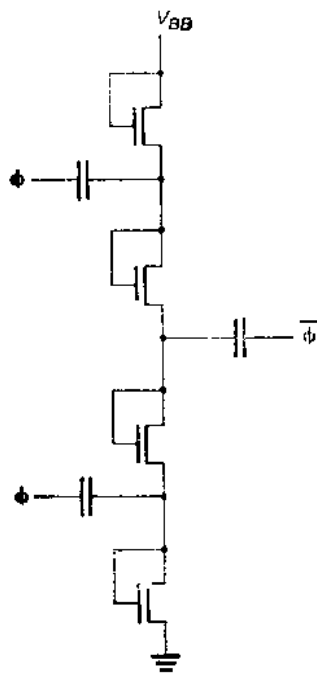


图5.67 一种图腾柱电路

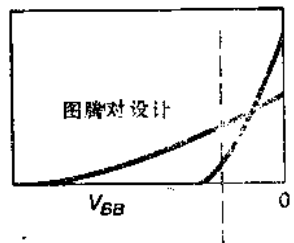


图5.68 具有相等电容的两种电路比较。在虚线处, 图腾柱的设计更好一些

已给出的电路可以总结归纳提供更低 V_{DD} 电压 [Dickson 76]。常用的图腾柱结构示于图 5.67 中。保持总的泵电容与以前的一样，可以对新的和老的设计作一比较。图 5.68 说明了这些器件的两条负载曲线。对于背栅电压接近 3V 的情况，可以看出，老的设计呈现出坏的电阻限制，而新的设计与之相反。采用新的设计，实际的 V_{DD} 电压是在反馈环路中被调整的，它使振荡器起振和停振。通常这是一个较好的设计，也可用图腾柱结构提供较高的正电压。

5.9 电路设计中的版图技术

电路设计师关心的主要版图技术是对信号线上的连线寄生电容和寄生电阻的控制，以及对电源总线电阻和电流负载能力的控制。在总体设计中，功能块布局和逻辑修改是用来控制信号寄生参量的两种主要技术。连线最短可以使芯片面积和寄生参量最小。平面布置图设计就是对 VLSI 芯片上信号流程、电源总线和逻辑功能块布局进行设计。平面布置图是为了进行版图设计，就象功能方块图是为了进行逻辑设计一样。

平面布置

一个 VLSI 电路通常是由一些功能块组成的。当这些功能块变换为硬件时，在功能块之间传输的信号将代表其整体布线。对这些功能块进行适当布局可以使它们之间的连线最短。在许多情况下，芯片上模块的几何位置分配是比较简单的。例如，微处理器中寄存器堆应当放在靠近算术逻辑单元 (ALU) 的地方。下一级的静态逻辑应当靠近、甚至就放在微码 PLA 之中。在另外一些情况下，寄存器或功能块的布局就不这么明显。指令寄存器 (IR) 是属于数据通道呢？还是属于芯片的控制部分呢？文字指令应当送到数据通道，而大多数其它指令则影响着机器的控制。在这些情况下，解决问题的办法就是计算一下连线多少。若把 IR 放在数据通道中，会有多少连线连到控制部分呢？而如果把 IR 放到控制部分中，情况又如何呢？

往往采用信号的局部译码来减少连线。当某些功能依赖于两个不同功能块中的信息时，就可以实现这一点。例如，考虑一个 ALU 的功能，它依赖于 ALU 的运算结果及下一条微指令码。我们可以把 ALU 的结果送到控制功能块，且用它决定下一条微指令，将该微指令译码再去控制 ALU。这是一种简捷的方法。一个更有效的方法是利用 ALU 的结果作为控制逻辑的输入，该控制逻辑电路实际上就连在 ALU 内部或者是在非常靠近 ALU 的地方。微码提供了一个“操作”指令，它控制着某个局部的控制端是否有效。这样作可以节省大量的连线并同时缩短临界通路的延迟时间。图 5.69 说明了一个使用这种技术的例子，这个例子是从寄存器堆中选择一个寄存器。在这个示例中，IR 是寄存器堆的一个部分，它含有一个字段，规定了哪个寄存器是被选中的。来自芯片控制部分的“SELECT REG”连线只告诉选择器是否要选择任何寄存器。局部逻辑决定了被选的寄存器。实际上微码并不知道哪个寄存器被选。如果有 31 个寄存器，则控制线的数目将从 5 减到 1。

另一个节省连线的方法是使逻辑双重化。例如，我们考虑一个如图 5.70 所示微处理器的微码 PLA 部分。PLA 的输出被锁存到微指令寄存器 (MIR) 中。MIR 控制着

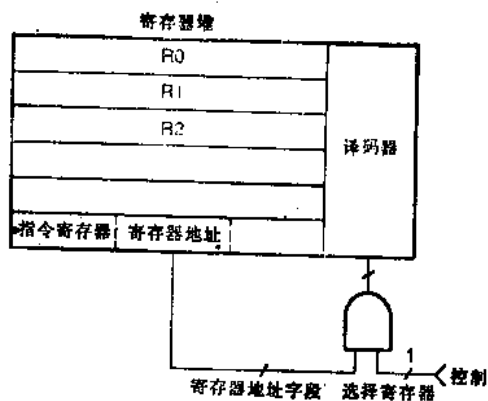


图5.69 带有译码器的寄存器堆。寄存器号码的选择是由局部逻辑完成的

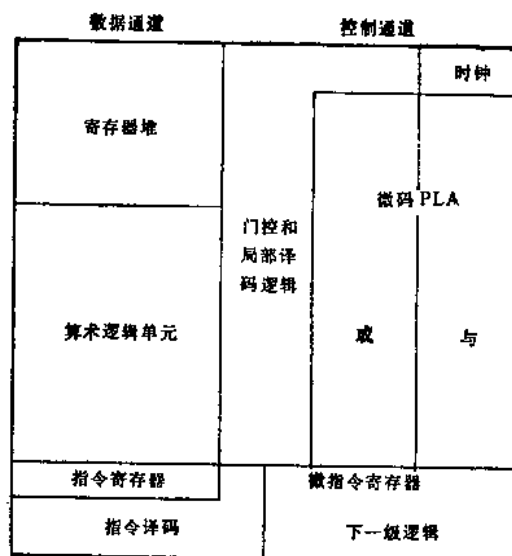


图5.70 简单微处理器的平面布置图

分散在数据通道中的许多功能。图 5.71 中给出了两个这样的控制点及其连线。从该图可以看出，其中一条连线必须上下穿过芯片。另一方面，已经有一条线基本上包含了同样的信息，它就在 PLA 的或矩阵中。图 5.72 给出了另外一种电路。该电路复制了 MIR 的功能，但使模块间的连线集中。

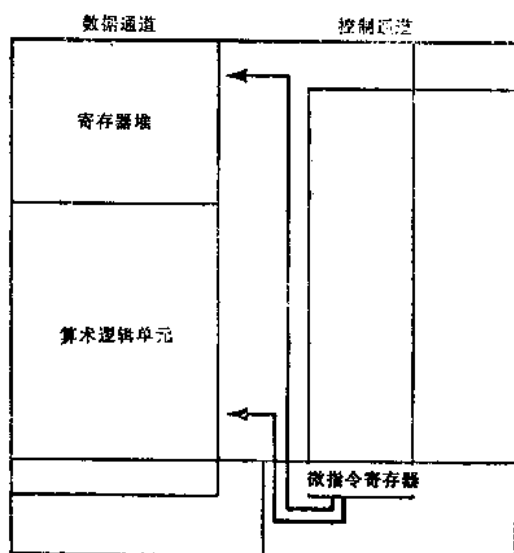


图5.71 数据通道控制线的布线

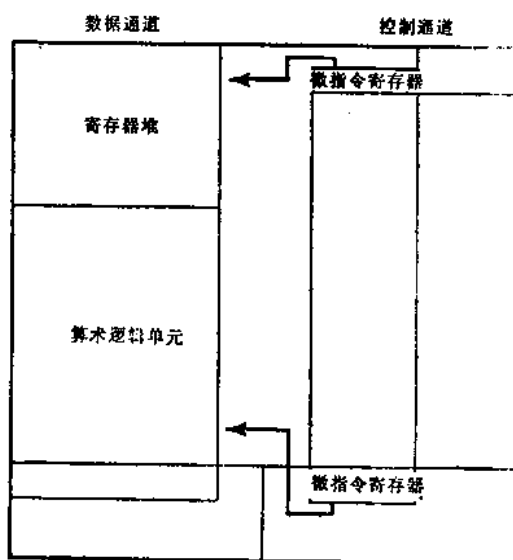


图5.72 具有冗余微指令寄存器数据通道控制线的较佳布线

芯片上的两种主要互连导体往往是互相垂直走线的。在我们已经介绍过的电路技术中，这两种导体是多晶硅和金属线。在更先进的工艺中，这两种导体可以是第一层金属线和第二层金属线。在许多情况下，数据流在一方向，而控制流在另一方向。究竟是数

据线还是控制线应该使用金属线,这是要进行折衷考虑的。控制往往是利用传输晶体管实现的。这就意味着数据是送入和送出到传输晶体管的扩散区的,而控制信号是通过多晶连到栅极的。这就启示我们,应当用多晶硅作为控制走线,用金属作为数据走线。在需要的时候,金属线必须向下连到扩散区去。图 5.73 说明了图 5.70 所示微处理器的主要金属信号布线。由于扩散区的高电容值,所以信号走线不能太长,这往往是较好的解决办法。但是,还存在一些问题。多晶硅具有高的电阻而且还要挂上许多门,使 RC 时间常数变得非常之差,从而引起波形畸变。如果阵列很宽,就有好几种可能的解决办法。其中之一是采用中继器 (repeater), 或者用把控制功能转换到金属线上的办法,从中心部位驱动矩阵。

具有双层金属时,情况稍有不同,但是其原理还是一样的。为了成为一个晶体管,这两层金属都必须与扩散区或多晶硅连通。第一层金属与第二层金属并不是等同的。在大多数具有双层金属的技术中,上层金属(第二层金属)必须与第一层金属相连,以便使它连接到多晶硅或扩散区上。在某些技术中,这些接触孔必须是错开的。这就是说,第一层金属与第二层金属的通孔不能处在第一层金属与多晶硅或扩散区接触孔的上方,即便是第二层金属可以直接与多晶硅或扩散区接触,在该接触的下边也不见得第一层金属连线可用的区域。

在所有这些情况下,制作第二层金属通孔要比第一层的接触孔费劲的多,因而合理的办法是使第二层金属通孔的数目减至最小。假如在数据通道中,要求连到数据线的接触孔比连到控制线的更多一些,则选择第一层金属作为数据线。在控制线中,多晶硅上的接触孔往往几个单元共用一个,所以这样作是有好处的。

功能框图向 VLSI 版图的转换并不是一种直接的翻译。功能框图主要强调逻辑功能而不是强调连线,它并不反映各模块的相对大小。图 5.74(a) 给出一个简单的数据通道方框图。在图 5.74(b) 中,给出了该功能框图的一种变换图。要注意所有被连线占据的面积。版图设计是很困难的,连线电容是很大的,面积随数据通道的宽度近似按平方律增长,这是因为它的长度和宽度同时在增大。图 5.74(c) 给出了它的 VLSI 格式版图。该版图是简捷的,信号通路很短,版图面积很小,并且随数据通道宽度的增加而线性增加。此外,很容易看出它是如何在水平方向上实现控制的,唯一的缺点是模块性。因而基本单元可以有一些连线在逻辑上并不“属于”这些单元,但却要从它们那里通过。

图 5.74(c) 中版图所反映的两个主要原则如下:其一,单元是邻近相接的而不是

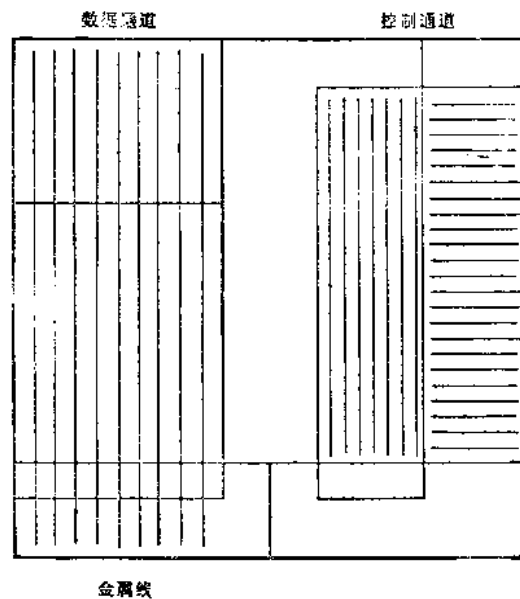


图 5.73 主要信号金属布线的平面设计图

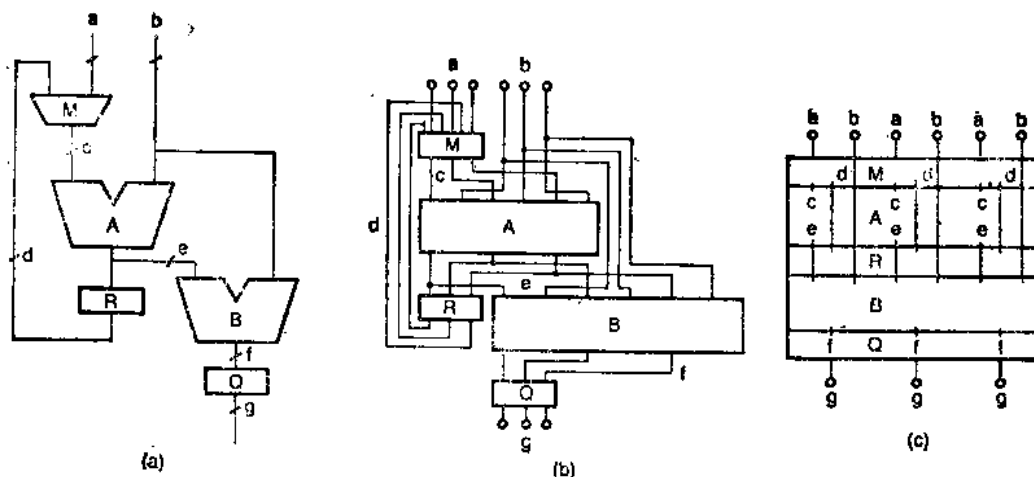


图5.74 将方块图转换到VLSI。(a)方块图，(b)向平面布置图的过渡；(c) VLSI式样的平面布置图

由长走线相连的。这一点可以改进性能并节省面积。为此，单元被延伸使间距匹配，如图 5.75 所示。第二个原则是避免首尾循环的走线，图 5.76 说明了这一原则。最好是穿过单元走线¹⁾而不是单元走线，对应于同一位的走线是交替的。

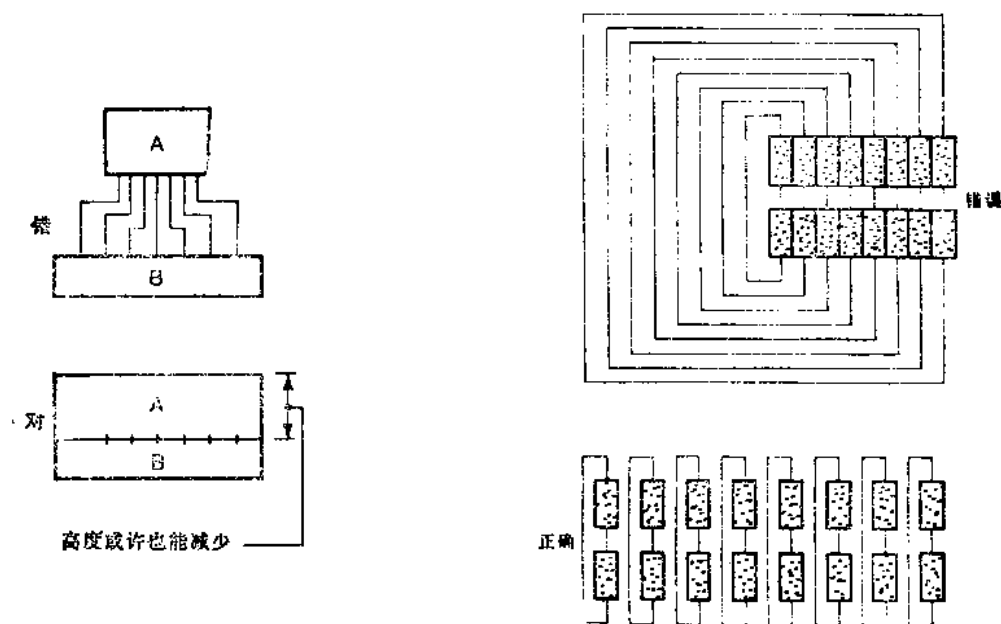


图5.75 实现模块之间连线常见的错误方法和正确方法

图5.76 在加入反馈链路的模块之间，连线中常见的错误方法和正确方法

在第一章中，我们曾看到为了适应工程变化，在 PLA 中设置备用行的概念。备用行的概念也适用于其它的版图设计领域。例如，在数据通道的设计中，为金属连线留下一些备用的位置就是一个很好的实例。在设计数据通道的过程中，就会发现这些备用部

1) 对于耦合噪声，必须小心从事。

位的用处。尽管这些严格构造的备用位置的使用也许是相当任意地取一块，砍一块，等等，但起码总的平面布置图是规则的是精心设计的。虽然不是所有备用位置都能用上，但还是改进了整体效率。这种技术可以扩展到芯片中其他更为随机的部分。不要把这些部分堆积得太密，以便于改变。

为了使两个信号相互作用，它们必须相互交叉。从平面布置的观点来看，如果一条信号线影响了另外一条信号线，重要的是它们在某处相遇。如果连线的一个矢量以某种任意的方式影响了另一个矢量，则必然存在全部“叉积”，其结果就表示为一种阵列结构。PLA，ROM，循环移位器等等都是这种简单原则的特例。阵列是一种容易改变的结构，可改变使得适应于安放它的版图区域。ROM的形状就是一个明显的例子。另外一个更为巧妙的例子可以在已经讨论过的微处理器中见到。我们确实希望从微码 PLA 或矩阵的一边取出某些信号，以便直接把它们馈送给条件逻辑电路。借助于阵列的稀疏性，可以用如图 5.77 中所示来实现这一点。虽然 PLA 不再是十分规整的了，但是整个芯片却可能更加规整。

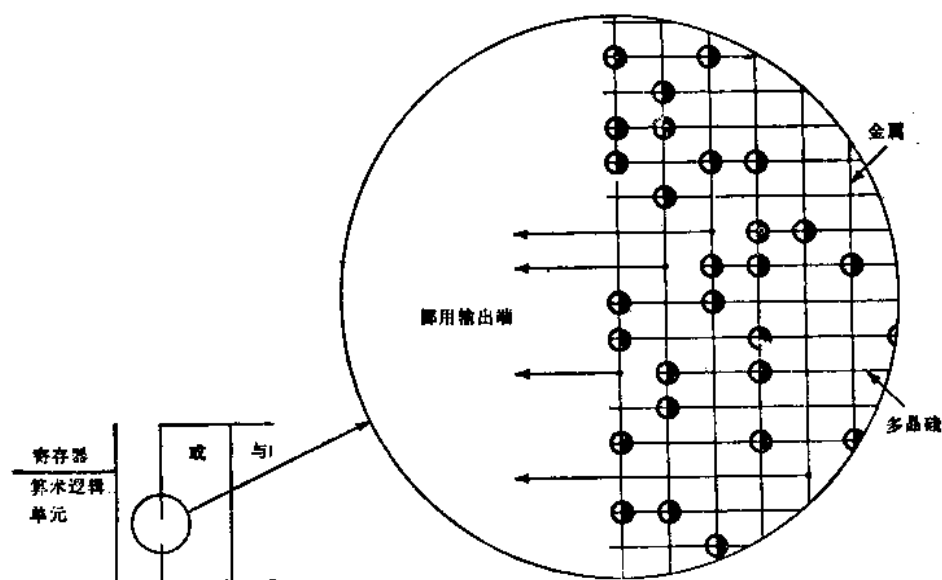


图 5.77 从 PLA 的一边挪用输出端

从平均的观点看，阵列越大越好。这样一来，设计诸如高性能读出放大器这类部件的成本就可以摊派到众多的晶体管上去。例如，倾向于把芯片内的寄存器放到阵列中去的原因之一是为了能共享公用读出放大器。

电源分配

在芯片平面布置时，最重要的问题之一是电源的布线。在单层金属工艺中， V_{DD} 和地线呈叉指式的树状布线，如图 5.78 所示，该图取自图 5.70 的微处理器电路。穿接是很困难的。在一定需要的时候，可以采用带有多重接触孔的低阻互连（有源区上的埋孔接触连接其上的多晶硅）。考虑穿接必须能通过 100mA 电流。在这种极端情况，比如说，低阻互连可能具有的最大电阻为 $10\Omega/\square$ ，因而每一方的低阻穿接就会产生 1 V 的压

降。由于电迁移的限制，在每一边就需要超过 50 个连到金属的 $2\mu\text{m}$ 接触孔。显然，对于密集地通过一个穿接来说， 100mA 是一个极大的电流。即便是 10mA 也是困难的，然而 10mA 仅相应于大约 20 个 nMOS 反相器。

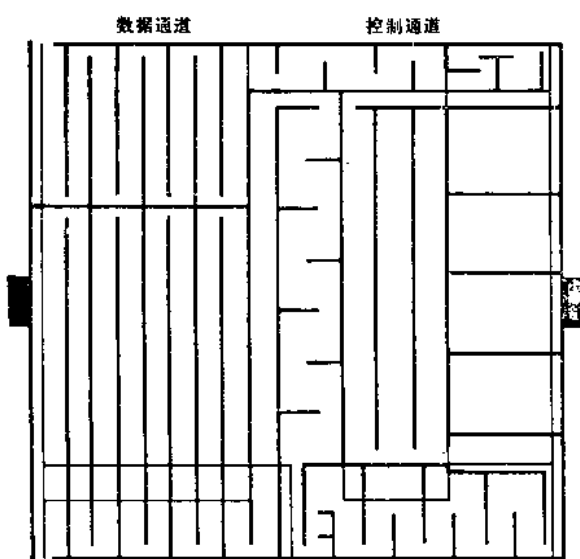


图 5.78 微处理器的电源和地布线。
电源线和地线呈叉指状的树型布线

而 10mA 仅相应于大约 20 个 nMOS 反相器。

电源线往往局部地分布在扩散区中，这是因为它总得要连到晶体管的源区和漏区。对于低功耗门，这种局部分配电源线并不坏，但对于高性能的门就必须十分小心。尽管这个问题很大，但是在采用双层金属时，很容易解决整个电源线的分配问题。

显然，当芯片的大部分同时处于瞬变过程时，对电源噪声来说，这时是最坏的条件之一。因此，一种方案是把芯片中预计会同时瞬变的电路的电源走线分开布线。这种情况的一个例子就出现在数据通道电路中。如果电源线横向水平地跨越位线，则会在

电源上引起大的脉动信号。但如果电源线沿信号线垂直走向，因沿数据通道的信息必然是错开的，则脉动电流就会非常之小。

芯片上电路的供电是高性能芯片的一个主要问题。任何地方的压焊丝均可具有 0.25 至 2nH 的电感（大约 $0.5\text{—}1\text{nH/mm}$ ）。 V_{DD} 线和地线往往是双线焊接（即在一个压焊块上压焊两条焊丝）。尽管这样作可以在某种程度上降低电感，但除非保持两条焊丝相互远离，否则电感不会像所期望的那样降低二倍，这是因为在两条线之间存在着互耦的缘故。按照惯例，在 40 脚双列直插电路封装中，往往电源引线是最长的一个，这就会使情况恶化。在电路封装中，电感可能是 $10\text{—}50\text{nH}$ [Prokop 78]。即使用一个外接的大电容，对于降低电源噪声电压也是徒劳的。常常用好几个 V_{DD} 和地线焊点，例如，因为驱动器存在极大的开关瞬变，且它们比内部电路能忍受更大的电源电压噪声，所以输出驱动器可以使用独立分开的电源管脚。

微单元版图设计中的电学问题

尽管大多数微单元版图设计问题超出了本书的范畴，但是由于它们影响着电路的性能¹⁾，所以某些版图设计问题还是适合于本书的题材。在第四章，我们讨论了有关扩散区中电源配线的噪声问题，以及噪声的电容耦合问题（特别是对于动态节点）。用户定制集成电路中的扩散区，最好刻成能降低源和漏电阻的形状。有时采用金属条来“恢复”地线或电源线。例如，在 PLA 中和 ROM 中，我们曾经这样用过。对于动态节点，最短沟长传输管的泄漏电流会成为一个问题。有时，为了消除短沟道效应，把这些晶体

1) 对版图设计技术比较认真的学生，版图设计方面的最好“书籍”是在硅中。对于版图设计和电路技术学习来说，一台好的显微镜是一件重要的工具。

管的沟道画得长一些。

为了简化芯片中要求不太严格电路部分的设计, 往往采用标准尺寸。这些器件的尺寸成为电路设计方法的一个部分, 即“在驱动一个 0.2pF 负载的时候, 采用尺寸为 A 的上拉器件”。这些器件尺寸是通过假定每个时钟节拍中含有某一数目的逻辑电平, 比如说 5 左右, 以及延迟在各级门之间等分而得到的。因此, 如果一个时钟节拍为 40ns , 则预先安排每个逻辑门的延时为 $6-8\text{ns}$ 。在采用特殊电路形式时, 也可以拟定相应准则, 即: “如果负载大于 0.5pF 则采用超驱动器”。借助于增加单元相似性以便重复使用, 以及帮助版图绘制者更快地学会给定芯片的最有效的版图设计习惯用语, 那么上述这些设计技术方法可以加快版图设计。

有时芯片要设计成只改变一层掩模 (往往是有源区或接触孔) 就可以对 PLA 重新编程。在要求具有这个自由度时, 经常会出现速度的损失, 而且还限于 PLA 或 ROM 的改变是芯片唯一的和必须的改变。在某些情况下这是有意义的, 但它并不是一般规律。

在含有埋接触的 nMOS 电路技术中, 标准的版图设计习惯是把晶体管安放在金属层之下。可以说, 各种 nMOS 版图设计的说明在本书中随处可见。

在 CMOS 电路技术中, 主要的考虑是阱面积的处理。有两种主要的版图设计选择。数据线的布线可以平行于或垂直于金属线的走向。图 5.79 给出一个简单的逻辑结构。在图 5.80 中, 所绘晶体管原理图是假定信息流平行于共用的金属电源线。在图 5.81 中, 上图被重画成另一结构, 其中假定数据流动方向垂直于金属电源线。第二种版图设计的习惯方式有几点好处, 它易于把阱区安放在边上而不是在中心处, 从抑制自锁效应的观点来看, 这一点是比较好的。还可以看出, 这个版图更加规整和紧凑, 这一点也会反映在最后的版图设计中。时钟线还是用金属层线。

我们曾在第二章和其它章节讨论过多晶硅线的延时问题。在此我们指出, 由于栅极的附加寄生效应和边缘电场的存在, 不把多晶硅线作成最小宽度, 有助于 RC 时间常数的改进。象任一寄生电路效应一样, 也可利用多晶硅线的延时作用。例如, 通过多晶硅线作输出驱动, 可以错开电流的脉冲从而降低脉冲尖峰噪声。

栅极很宽的晶体管的版图设计细节也会对 VLSI 芯片的性能和面积有很大的影响。大个晶体管往往出现在压焊块和时钟驱动器的末级。有两个重要的问题。其一, 是要考虑寄生电容; 其二, 就是我们常说的多晶硅电阻。我们曾看到, 扩散区的边缘电容是逻辑门的主要电容负载, 这是由于场区注入的缘故。当确实如此时, 为了方便, 我们把晶体管边界处的扩散区与多晶硅的交界线看作是扩散区边缘电容。实际上这个区域具有不同的且往往是更小的电容值。当然, 即使不是这样, 减小扩散区面积也是有用的。从所有这些观点来看, 图 5.82 中所示的封闭或环形晶体管是一种很好的结构。然而, 它也有一些不利之处。这种结构的拐角区几乎没有提供载流能力, 但是该区却对门的输入

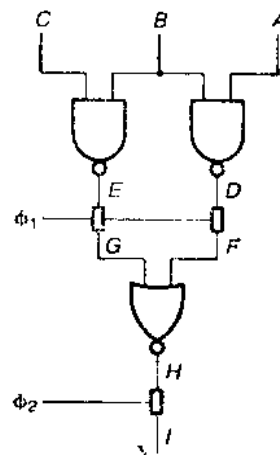


图 5.79 逻辑门的集合

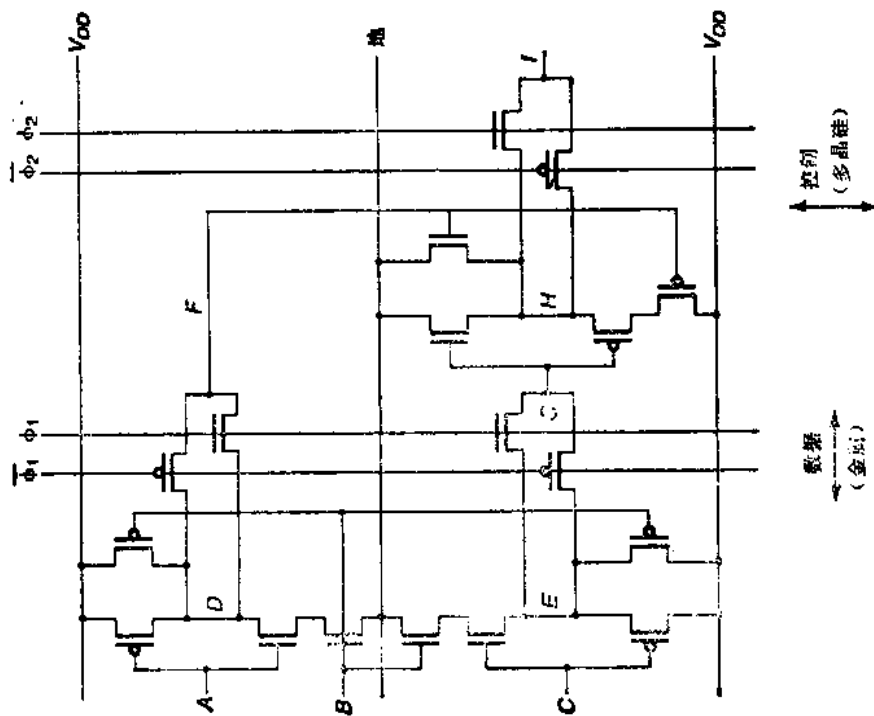


图 5.80 实现图 5.79 逻辑电路的一种拓扑结构

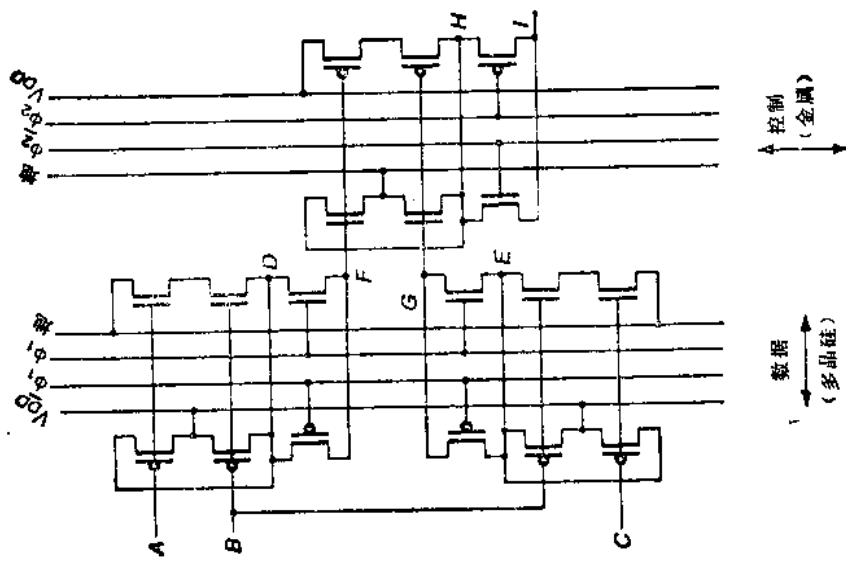


图 5.81 实现图 5.79 逻辑电路的另一种拓扑结构。时钟线用金属线，并很好地利用了断的边界区，很容易在靠近断的边界区安放输入结构

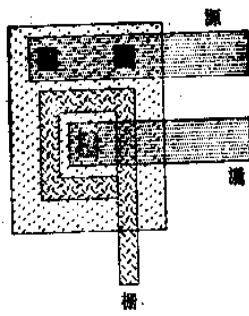
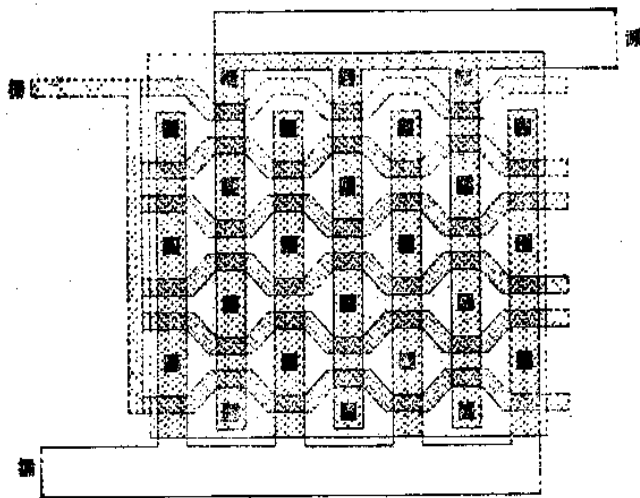
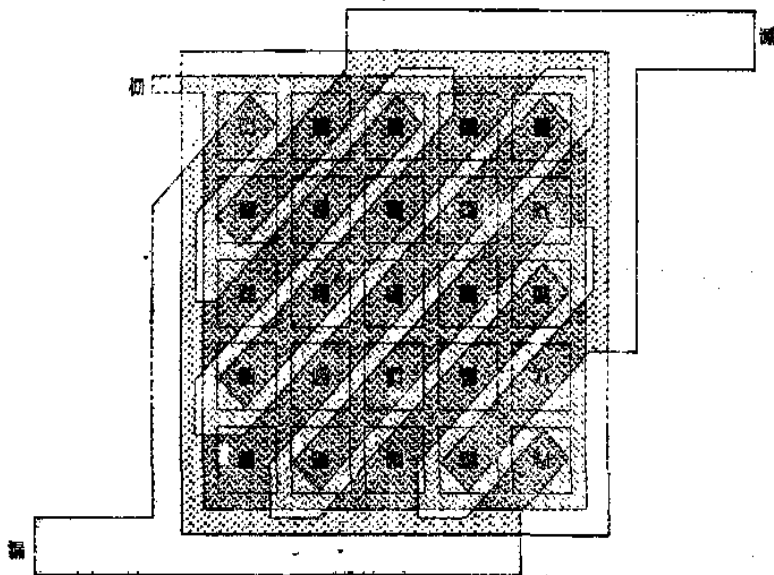


图5.82 环形下拉晶体管



(a)



(b)

图5.83 充分地考虑了 RC 时间常数问题的栅极很宽的晶体管
(a) 把等位节点短路了；(b) 采用了“华夫饼干”模式的设计

电容有所贡献。在某些具有高电源电压与沟道长度比的工艺中，在这些拐角区中的高电场甚至会引起器件击穿。在这种结构中，外环扩散区的串联电阻也是必须加以考虑的。

在高性能 VLSI 中，压焊块驱动晶体管很可能具有几毫米的栅宽¹⁾。在这种情况下，扩散区和多晶硅线的电阻是非常严格的。图 5.83 给出两种结构用来构成栅极非常宽的晶体管。将一些小段硅栅并联起来可使多晶硅的电阻最小。用金属条连接扩散区，可使扩散区电阻保持较低值。图 5.83 中给出单个晶体管的版图设计结果。若有许多个晶体管，则按照适合于手头现有的实际工艺，或者一个换一个地安放或者交叉地安放。

有时，在 MOS 数字电路设计中需要对电路元件进行匹配。关键在于要对这些元件匹配到既使在对准不良的情况下，也能保持良好匹配。图 8.54 说明了某些匹配的和不匹配的晶体管。在许多情况下，不仅晶体管匹配是重要的，而且电容和电阻的匹配也是重要的。图 5.85 表示了两个电容，其中之一是另一个的 2 倍。电容的面积和边界都按比例变化也是同样重要的²⁾。作为一级近似，边长按比例变化意味着要小心地作到按照边缘形状均匀地收缩或胀大。时钟读出放大器（见 5.5 节）是一个需要电阻匹配、电容匹配和驱动能力匹配的电路。图 5.86 给出一个不紧凑的但是平衡的交叉耦合电路。

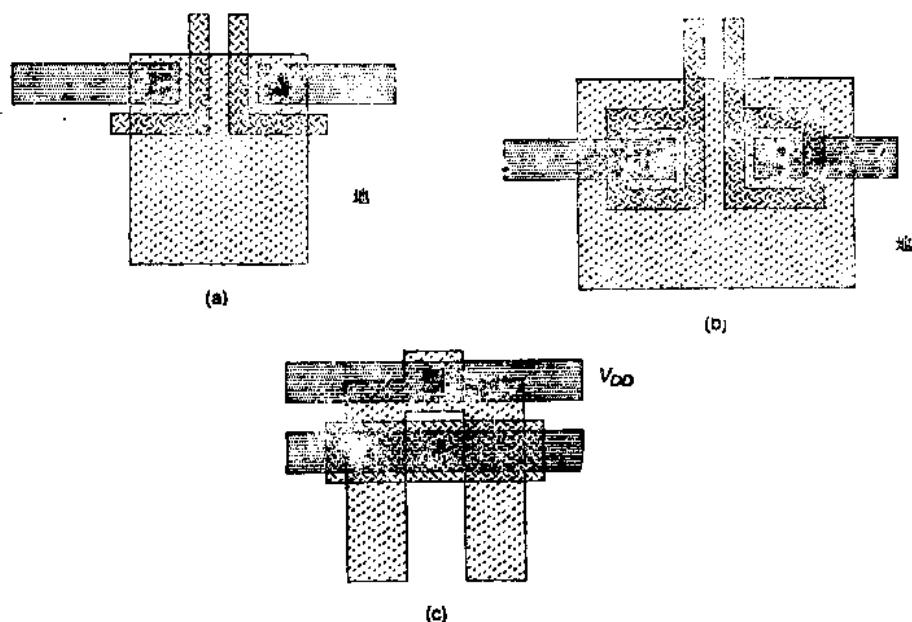


图 5.84 晶体管对。(a) 在对准不良时导致不对称，(b) 和 (c) 在对准不良时仍保持对称性



图 5.85 面积和边界尺寸都匹配的电容

- 1) 在 CMOS 中，往往采用射极跟随器作为输出驱动上拉管。
- 2) 这些只是最初步的考虑。在 MOS 模拟电路的设计中，还要确定一下匹配角的数目，而且要把电容 $2C$ 的一半放到小电容的一侧；另一半放到另外一侧。这是为了补偿氧化层厚度的线性梯度。

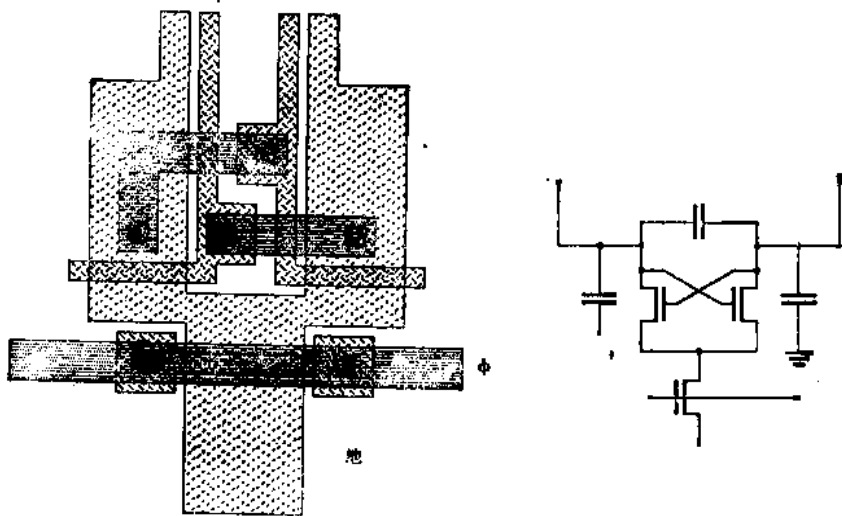


图5.86 在对中不良的情况下驱动和负载都对称的电路

5.10 小 结

在本章中，讨论了电路的优化。对给定的关键通路，电路优化技术一般有下列六项：

1. 直接的方法是改变图中各个管子的栅宽和栅长，而且对版图设计影响最小（第5.1节）。
2. 在高阻抗源和低阻抗负载之间加入缓冲级，缓冲级的数目有最佳值（第5.2节）。
3. 在大扇出的节点处要对各负载去耦，而且要分别使每条最终的通路最佳化。
4. 为了利用驱动能力很强的系统时钟的低阻抗特性，采用了时钟缓冲器。这种技术只能用在具有合适时钟沿的情况中。在芯片的结构设计中，时钟电路技术的用处是激发多重时钟沿（5.2节）。
5. 在nMOS电路中，对大电容节点进行预充，既可以得到较好的速度性能又可得到较低的静态功耗。在输出上升瞬变延迟和下降瞬变延迟明显不同时，预充可以改进MOS电路的速度性能。如同上述第4项技术一样，预充要求适当的时钟沿。如果系统结构许可的话，有时仍然可以用定时链产生。
6. 引用一些特殊的电路。这些电路对于高扇入或高扇出情况特别有用。比如像单端读出放大器和超驱动器，它们与普通的技术是不同的（5.2节到5.4节）。

除了这些属于电路范畴的技术外，在其它一些抽象范畴中也可以实现优化。这些范畴包括整机结构、逻辑和版图设计。

特殊情况产生于要求特殊电路的芯片设计。这些情况往往出现在芯片的输入部分（输入保护，施密特触发器，电平转换电路）和检出部分（检出驱动器，电压参数发生器）。为有效利用大型的规则结构，往往需要特殊的接口电路。

习 题

- 5.1 分析图 5.87 中的读出放大器。它的优点是什么？版图设计的关键问题是什么？讨论输出对于器件阈值匹配的灵敏度[Furuyama 81]。

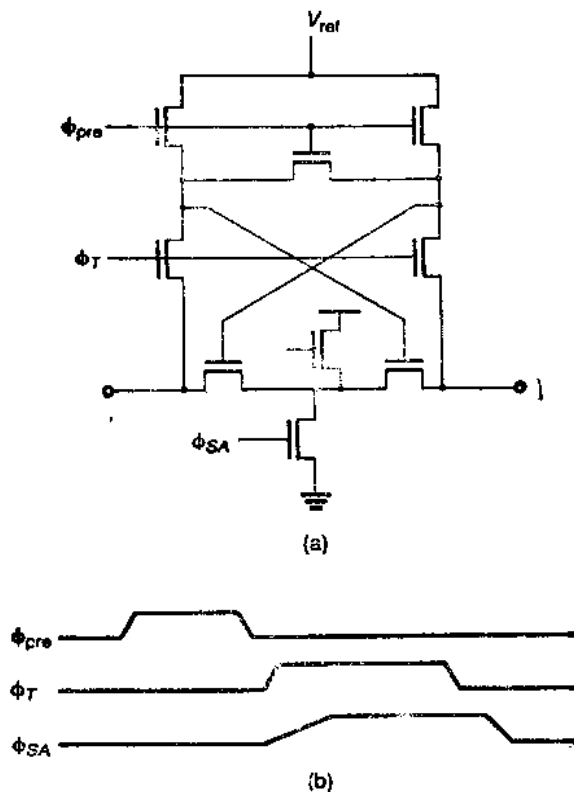


图 5.87



图 5.88

- 5.2 在图 5.88 中给出了称之为“λ 二极管”的元件，通过该两端元件的峰值电流是多少？
- 5.3 设计图 5.89 黑方块中的电路，使之适应相应位线并尽可能快地检测出读出信号。采用 $2\mu\text{m}$ nMOS 工艺，电路最坏情况下的功耗不要超过 1mW ，使用多少时钟可根据你的需要决定。运行 SS 模拟程序以便给出存取时间，运行 FF 模拟程序以便给出功耗值，运行 FS 和 SF 程序以证明其功能。假定时钟不随工艺加工而变。通过读“1”，再读“0”，再读“1”来验证你的设计。并画一个粗略版图以提取寄生参数。△
- 5.4 图 5.90 中的电路可用以对一差分输入信号进行电平移动。通过分析和 SPICE 模拟，研究该电路并报告你的发现。反相器的微分增益起什么作用？再看图 5.91 中的电路，为什么这个电路具有较好的性能？△
- 5.5 若有一个由相同 CMOS 反相器组成的无穷长链，为了达到最小的对延迟，试证明上拉器件与下拉器件的最佳宽度比为

$$\frac{W_{pn}}{W_{pd}} = \sqrt{\frac{\mu_n}{\mu_p}}$$

假设器件的有效沟道长度都一样，并忽略互联电容。

- 5.6 采用示例中给出的工艺之一，设计一个 14.5MHz 的晶体振荡器。晶体振荡器的等效电路示于图 5.92 中。用 $C_0 = 5.2\text{pF}$ ， $C = 0.027\text{pF}$ ， $L = 4.4\text{mH}$ ， $10 \leq R_1 \leq 25$ ， $50 \leq R_2 \leq 100$ 。试确认该振荡器以基本模式起动，在图 5.93 中给出了所建议的电路图。△
- 5.7 根据给定触发点的输入规范、负载电容和最坏情况下的上升下降时间，写一个设计 CMOS 施密特触发器的计算机程序。这一个程序表示了程序性的电路设计。把你的预期结果与 SPICE 模拟结果进行比较。你认为这样一个方法对工艺能有多大独立性？△

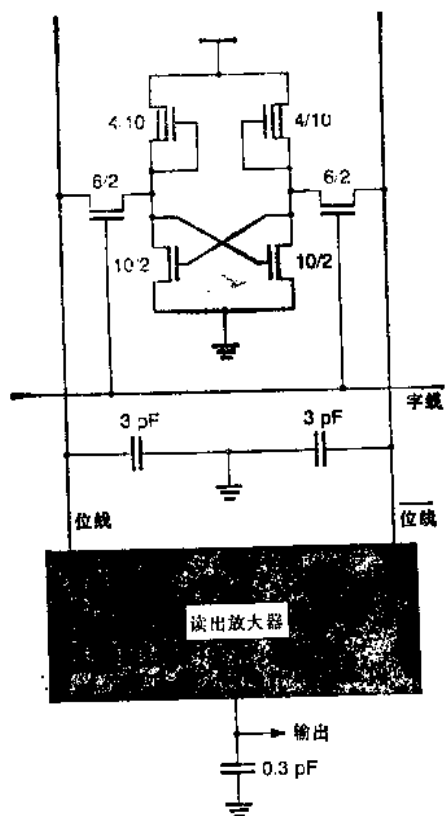


图 5.89

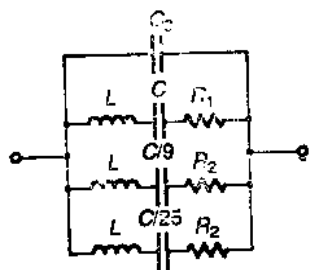


图 5.92

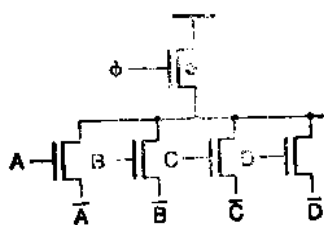


图 5.94

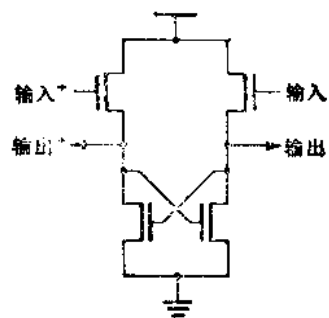


图 5.90

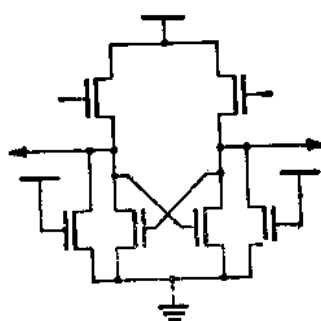


图 5.91

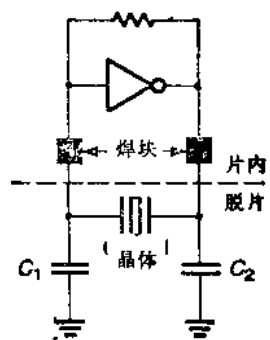


图 5.93

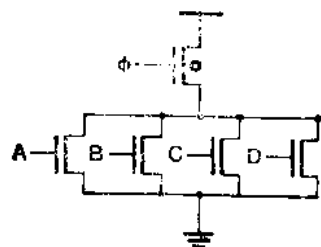


图 5.95

- 5.8 已经肯定, 与图 5.95 中的电路相比, 图 5.94 中的预充或非门对短沟道效应更不灵敏。如果这个断言是真实的, 为什么?[White 80]
- 5.9 对图 5.96 中所示电路, 推导一个简单的 $I-V$ 特性表达式。

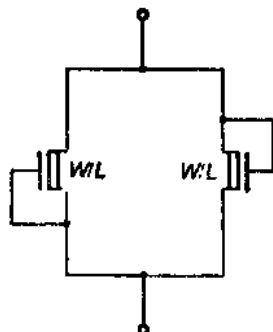


图 5.96

- 5.10 自举驱动器依赖于在下拉管断开之前, 自举电容是否被很好地充电。在驱动器中, 我们早已看到, 电容充电时间和延时之间极少相关。设计一个自举驱动器, 使其中的这些时间能更好地相关。在所有工艺的各个小范围内对你的设计进行模拟。
- 5.11 在图 5.22 的自举驱动器中, 当假定 v_2 超过 V_{DD} 时, 存在着 M_4 的电荷将会泄漏的潜在问题。假定已有时钟波形 ϕ , 设计一个电荷泵电路, 使得 v_2 较高时能够保持住。
- 5.12 M_2 的电阻对图 5.22 电路是一个限制。设计并模拟一个驱动器, 其中引用一个附加级去控制 M_2 的开关。(允许管子栅级更宽一些)。假定负载为 100pF 。
- 5.13 连接到 MOS 芯片的总线, 通过具有 100Ω 的外部上拉管连到 5V 电源上。通过单个芯片上源端接地增强型晶体管的漏端将每个芯片连到母线上。要求最小的电压摆幅为 1V (把总线下拉到 4V)。摆幅大于这个值则耗散过量功率, 并且产生附加噪声。设计一个电路限制输出管的栅电压, 以便控制检出摆幅, 并具有合适的工艺稳定度。设计时要小心, 使得电路不会限制最坏情况下驱动器的速度。
- 5.14 试解释如何利用电荷泵实现电平移动。对于输入电压接近或低于 V_{TE} 的情况, 这个能力是很临界的。

参 考 文 献

- [Arai 78] E. Arai and N. Jeda, "A 64-kbit Dynamic MOS RAM," *IEEE J. Solid-State Circuits* SC-13:333—338, 1978.
- [Bandler 79] J. W. Bandler and M. R. M. Rizk, "Optimization of Electrical Circuits," *Engineering Optimization, Mathematical Programming Study 11*, North-Holland, Amsterdam, 1979.
- [Barnes 80] J. J. Barnes and J. Y. Chan, "A High Performance Sense Amplifier for a 5V Dynamic RAM," *IEEE J. Solid-State Circuits* SC-15:831—838, 1980.
- [Blaser 78] E. M. Blaser and D. A. Conrad, "FET Logic Configuration," *IEEE International Solid-State Circuits Conf.*, 14—15, San Francisco, Calif., 1978. (Cascode-based logic gates)
- [Brayton 80] R. K. Brayton and R. Spence, *Sensitivity and Optimization, Computer-Aided Design of Electronic Circuits*, 2, Elsevier, New York, 1980.
- [Carr 72] W. N. Carr and J. P. Mize, *MOS/LSI Design and Applications*, McGraw-Hill, New York, 1972.
- [Chan 80] J. Y. Chan, J. J. Barnes, C. Y. Wang, J. M. DeBlasi, and M. R. Guidry, "A 100ns 5V Only $64\text{K} \times 1$ MOS Dynamic RAM," *IEEE J. Solid-State Circuits* SC-15:839—846, 1980.
- [Cheng 80] E. K. Cheng and S. Domenik, "Substrate Bias Generation Design," *Lambda*, 1:74—75,

1980.

- [Cooper 77] J. A. Cooper, J. A. Copeland, and R. H. Krambeck, "A CMOS Microprocessor for Telecommunications Applications," *IEEE International Solid-State Circuits Conf.*, 138—139, Philadelphia, Pa., 1977. (pseudo-nMOS)
- [Crawford 67] R. H. Crawford, *MOSFET in Circuit Design*, McGraw-Hill, New York, 1967.
- [Dickson 76] J. F. Dickson, "On-Chip High-Voltage Generation in nMOS Integrated Circuits Using an Improved Voltage Multiplier Technique," *IEEE J. Solid-State Circuits* SC-11, 374—378, 1976.
- [Dingwall 77] A. G. F. Dingwall, R. E. Stricker, and J. O. Sinniger, "A High-Speed Bulk C²L Microprocessor," *IEEE International Solid-State Circuits Conf.*, 136—137, Philadelphia, Pa., 1977.
- [Elmasry 81] M. I. Elmasry, *Digital MOS Integrated Circuits*, IEEE Press, New York, 1981.
- [Furuyama 81] T. Furuyama, S. Saito, and S. Fujii, "A New Sense Amplifier Technique of VLSI Dynamic RAMs," *IEEE International Electron Device Meeting*, 44—47, Washington, D. C., 1981.
- [Gallace 77] L. J. Gallace and H. J. Pujol, "The Evolution of CMOS Static-Charge Protection Networks and Failure Mechanisms Associated with Overstress Conditions as Related to Device Life," *Proc. IEEE International Reliability Physics Symposium*, 149—157, Las Vegas, Nev., 1977.
- [Glasser 84] L. A. Glasser and L. P. J. Hoyte, "Delay and Power Optimization in VLSI Circuits," *21st Design Automation Conf.*, 529—535, Albuquerque, N. M., 1984.
- [Heller 76] L. G. Heller, D. P. Spampinato, and Y. L. Yao, "High-Sensitivity Charge-Transfer Sense Amplifier," *IEEE J. Solid-State Circuits* SC-11, 598—601, 1976.
- [Heller 79] L. G. Heller, "Cross-Coupled Charge-Transfer Sense Amplifier," *IEEE International Solid-State Circuits Conf.*, 20—21, New York, 1979.
- [Huffman 77] D. Huffman, B. Green, and D. Segers, "Minimizing Threshold Voltage Temperature Degradation with a Substrate Bias Generator," *MOSTEK Applications Information*, 1977.
- [Ipri 77] A. C. Ipri, "Lambda Diodes Utilizing an Enhancement Depletion CMOS/SOS Process," *IEEE Trans. Electron Devices* ED-24, 751—756, 1977.
- [Kang 81] S. Kang, "A Design of CMOS Polycells for LSI Circuits," *IEEE Trans. Circuits and Systems* CAS-28, 838—943, 1981.
- [Knepper 78] R. W. Knepper, "Dynamic Depletion-mode, An E/D MOSFET Circuit Method for Improved Performance," *IEEE J. Solid-State Circuits* SC-13, 542—548, 1978.
- [Koomen 72] J. Koomen and J. van de Akker, "A MOST Inverter with Improved Switching Speed," *IEEE J. Solid-State Circuits* SC-7, 231—237, 1972. (Bootstrapped superbuffers)
- [Lee 79] J. M. Lee, J. R. Breivogel, R. Kunita, and C. Webb, "A 80ns 5V-Only Dynamic RAM," *IEEE International Solid-State Circuits Conf.*, 142—143, New York, 1979. (V_{DD} generation)
- [Lin 75] H. C. Lin and L. W. Linholm, "An Optimized Output Stage for MOS Integrated Circuits," *IEEE J. Solid-State Circuits* SC-10, 106—109, 1975.
- [Lynch 74] W. T. Lynch and H. J. Boll, "Optimization of the Latching Pulse for Dynamic Flip-Flop Sensors," *IEEE J. Solid-State Circuits* SC-9, 49—55, 1974.
- [Martino 80] W. L. Martino, Jr., J. D. Moench, A. R. Borman, and R. C. Tesch, "An On-Chip Back-Bias Generator for MOS Dynamic Memory," *IEEE J. Solid-State Circuits* SC-15, 820—825, 1980.
- [Millman 79] J. Millman, *Microelectronics*, McGraw-Hill, New York, 1979.
- [Matson 83] M. Matson, *Circuit Level Optimization of Digital MOS VLSI Designs*, Ph. D. Thesis Proposal, Massachusetts Institute of Technology, Cambridge, Mass., June 1982.
- [Mohsen 79] A. M. Mohsen and C. A. Mead, "Delay-Time Optimization for Driving and Sensing of Signals on High-Capacitance Paths of VLSI Systems," *IEEE Trans. Electron Devices* ED-26, 540—548, 1979.
- [Pashley 76] R. D. Pashley and G. A. McCormick, "A 70 ns 1 K MOS RAM," *IEEE Internati-*

- onal Solid-State Circuits Conf.: 138—139, 238, 1978.
- [Penney 72] W. M. Penney and L. Lau, *MOS Integrated Circuits*, Van Nostrand, Philadelphia, Pa., 1972. Reprinted in 1979 by Robert E. Krieger, New York.
- [Proebsting 75] R. J. Proebsting, "Low Power, High Speed, High Output Voltage FET Delay-Inverter Stage," U. S. Patent 3,898,479, August 1978.
- [Prokop 78] J. S. Prokop and D. W. Williams, "Chip Carriers as a Means for High-Density Packaging," *IEEE Trans. Components, Hybrids, and Manufacturing Technology* CHMT-1, 297—304, 1978.
- [Puri 78] Y. Puri, "Substrate Voltage Bounce in nMOS Self-Biased Substrates," *IEEE J. Solid-State Circuits* SC-13, 515—519, 1978.
- [Schmidt 65] J. D. Schmidt, "Integrated MOS Random-Access Memory," *Solid-State Design*, 21—25, 1965. (sRAM)
- [Schroeder 77] P. R. Schroeder and R. J. Proebsting, "Clock Generator and Delay Stage," U. S. Patent 4,061,933, December, 1977.
- [Stein 72] K.-U. Stein, A. Sihling, and E. Doering, "Storage and Sense/Refresh Circuit for Single-Transistor Memory Cells," *IEEE J. Solid-State Circuits* SC-7, 336—340, 1972.
- [Tokuda 83] T. Tokuda, K. Okazaki, K. Sakashita, I. Ohkura, and T. Enomoto, "Delay-Time Modeling for ED MOS Logic LSI," *IEEE Trans. Computer-Aided Design* CAD-2, 129—134, 1983.
- [Troutman 83] R. R. Troutman, "Epitaxial Layer Enhancement of n-Well Guard Rings for CMOS Circuits," *IEEE Trans. Electron Devices Lett.* EDL-4, 438—440, 1983.
- [Vergnieres 80] B. Vergnieres, "Macro Generation Algorithms for LSI Custom Chip Design," *IBM J. Res. Develop.* 24, 612—620, 1980.
- [White 80] L. S. White, Jr., N. H. Hong, and D. J. Redwine, "A 5 V-Only 64K Dynamic RAM," *IEEE International Solid-State Circuits Conf.*: 230—231, 1980.
- [Wooley 80] B. A. Wooley, "Course notes for EECS 243," University of California, Berkeley, 1980, unpublished.
- [Yee 78] Y. S. Yee, L. M. Terman, and L. G. Heller, "A 1 mV MOS Comparator," *IEEE J. Solid-State Circuits* SC-13, 294—297, 1978.

第六章 时钟与通讯

当一个信号影响了另一个信号时，它们一定不仅在空间上，而且在时间上产生相互作用。在 VLSI 系统中，要安排好成千上万个信号以便每一部分数据在预定的时间到达指定的地点，这是一项极其复杂的任务，对于高性能设计更是如此。

解决这个复杂问题的方法之一是采用时钟。这种方法使数字系统中有序而受控的信号流同步，又由于限制了有效的逻辑形式而使电路设计的复杂性降低。后者和我们在第一章中的逻辑门设计方法一样，用限制可能采用的电路形式来简化设计过程。

对于数字电路，我们必须把逻辑值及时间两者复原，这就是说，必须搞清在什么情况下用一个模拟电压波形来代表某个逻辑值是正确的。我们知道，当信号通过网络时，它们都将产生无法精确预期的延迟（譬

如说，要想使一台周期时间为 200ns 的计算机运行一年而又不时钟或其它再同步方法，则必然要求计算机中所有延迟时间的精度优于 $6 \times 10^{-23}\text{s}$ ）。

由于需要把信号合拍地汇集在一起，这就必须开发一种处理延迟不确定性的技术。一种称为“时钟定时”的方法是利用周期地阻滞所有信号而使全部延迟都相等，也就是使它们都同等地差；另一种方法是使所有信号在最慢的一个到达前都等待着，这种方法叫做“自定时信号”。采用时钟来及时复原信号是更为通用的方法。

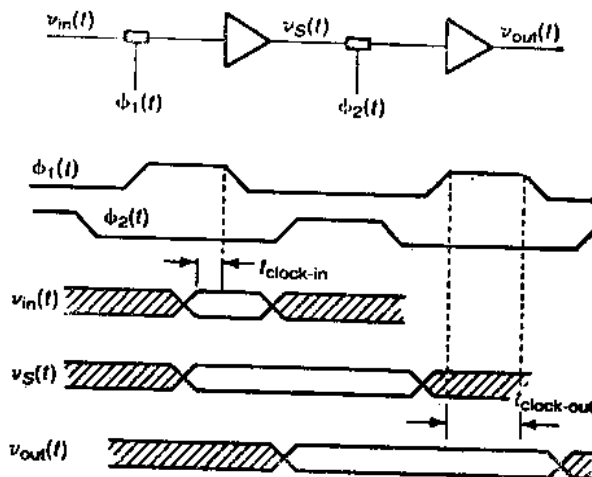


图 6.1 用时钟定时复原时间

图 6.1 是信号 $v_{in}(t)$ 通过一组由非重叠时钟控制的动态寄存器的波形，信号在从 v_{in} 到 v_S 再到 v_{out} 的同时，既被延迟也被复原。毫无疑问，对 $\phi_1(t)$ 和 $v_{in}(t)$ 都有效的 $t_{\text{clock-in}}$ 不应当减小至 0（为简单计，我们假设寄存器响应是瞬时的），把电源加到逻辑门上不过使电压得以复原，而及时复原信号的唯一方法是使它延迟。

6.1 单相时钟定时

数字系统中用时钟来阻滞信号，使它恰好在预定的时间开始向下一个逻辑级传送，同时寄存器和时钟配合使用就可以使信号在开始传送前贮存在某个地方。

寄存器和时钟可以用 TTL 组成，一般采用边缘触发的触发器结构，有时候也用 MOS，寄存器大多为传输晶体管。图 6.2 为一个单相有限状态机，用它作为说明这种系统的一个实例。在所有时钟方法中，单相时钟定时使用的晶体管最少，但却因组合逻辑块中

逻辑元件的速度上下都受到限制而包含复杂性。设 $T_L + T_H = T_P$ ，其中 T_P 为时钟的周期， T_H 和 T_L 分别是时钟为高电平和低电平时的时间。信号通过组合逻辑时根据它们逻辑值的

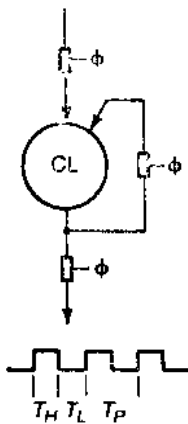


图6.2 单相时钟定时 $T_H + T_L = T_P$ ，组合逻辑一定不能太快，也不能太慢

不同而有不同的速度，我们分析两种极端情况。首先分析最慢的有效逻辑信号通过逻辑块需要的时间，如果我们要使时钟定时与数据无关，则最长的传播延迟必须小于 T_P ；其次分析信号（甚至于可能是由于内部竞争冒险产生的尖峰所造成的假信号）到达逻辑块输出端可能取的最短时间，这个时间必须大于 T_H 。令 τ_{CL} 代表组合逻辑块的延迟范围，即从最短的无效延迟到最长的有效延迟，我们得到

$$T_H < \tau_{CL} < T_P \quad (6.1)$$

上式表明，信号通过组合逻辑块的每一个延迟都必须大于 T_H 并小于 T_P 。正是这种双边约束特性使单相时钟定时难以实现。我们当然需要关注最慢的通道，不仅如此

6.2 两相时钟定时

将图 6.2 作一简单的修改就取消了双边约束，如图 6.3 所示。这里用的是两相非重叠时钟，时钟的波形示于图 6.4， T_1 和 T_3 分别是 ϕ_1 和 ϕ_2 为高电平时的时间， T_2 为 ϕ_1 到 ϕ_2 之间电平为低时的时间， T_4 则是 ϕ_2 到 ϕ_1 之间的低电平时间。利用前面用过的分析方法，得：

$$\tau_{CL} < T_1 + T_3 + T_4 \quad (6.2)$$

或

$$\tau_{CL} < T_P - T_2 \quad (6.3)$$

其中

$$T_P \equiv T_1 + T_2 + T_3 + T_4 \quad (6.4)$$

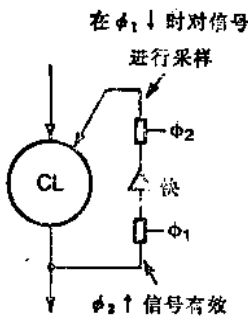


图6.3 两相时钟定时电路解决了双边约束的困难

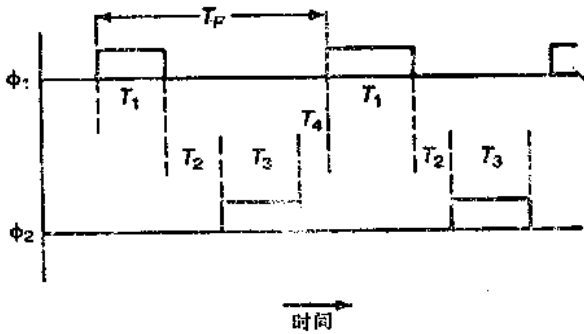


图6.4 两相时钟定时的非重叠波形图

ϕ_2 变高时信号开始通过组合逻辑块传输，并且必须在 ϕ_1 变低之前结束。值得注意的是，我们把双边约束(6.1)式已简化为单边约束(6.3)式了，无论是有效信号或是无效

信号, 都可以以任意快的速度通过组合逻辑块而不会造成竞争的条件。为了达到这一目的, 我们付出了三个方面的代价: 锁存¹⁾电路从一个传输门增加为两个传输门加上一个缓冲器; 要产生更为复杂的时钟; 芯片周围的时钟线增加很多。由于研究的是VLSI, 从设计的角度来说, 增加时钟线比增加晶体管难度更大, 但是, 即使时钟线增加了一倍, 克服设计困难所付出的代价也是不大的。所以, 像微处理器这种常规电路, 采用的时钟已多达8个。

有了两相时钟就多了一个附加的自由度。我们从式(6.3)可以看出, 在每一个周期中 T_2 是浪费掉的, 但在某些情况下, 这部分时间能够争取回来。假设我们可以把组合逻辑分成两组, 如图 6.5 所示, 通过这两组逻辑模块的延迟为一种两维约束:

$$\tau_C < T_1 + T_3 + T_4 \quad (6.5)$$

$$\tau_L < T_1 + T_2 + T_3 \quad (6.6)$$

且

$$\tau_L + \tau_C < T_P \quad (6.7)$$

任何满足所有这些约束条件的 τ_L 和 τ_C 的组合都是允许的, 解的空间示于图 6.6。可以看出, 如果 τ_L 大则 τ_C 一定要小, 反之亦然。所以这时能够在模块之间自由地来回调整延迟条件而不改变时钟波形, 这正是非边缘触发系统的特征, 尽管这样做会增加设计的复杂性, 有时候还是要利用它。系统的总延迟 $\tau_{CL} = \tau_C + \tau_L$, 它可以接近于 T_P 。

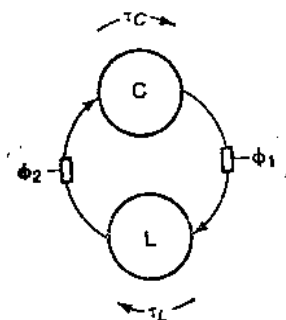


图 6.5 把组合逻辑分成两部分, 在这两个模块之间如何分配总延迟有相当大的自由度

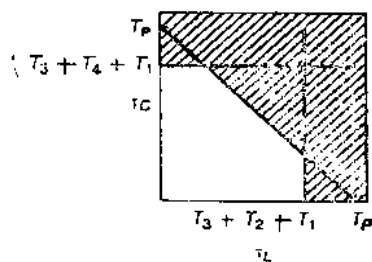


图 6.6 两相非重叠时钟的约束空间。可以用 C 的延迟换取 L 的延迟

图 6.7 为多级组合逻辑块, 其中每一反馈通道都同时被 ϕ_1 和 ϕ_2 控制的旁路门切断。此图用简单的方法说明了每一个组合逻辑块都具有这样的特点, 即所有输入端用某一相时钟断开, 所有输出端则用另一相时钟断开。式(6.5)、(6.6)给出的约束条件对每一逻辑块仍然有效, 但现在式(6.7)的约束条件却十分复杂了, 因此, 常常采用 $\tau_C < T_1$ 和 $\tau_L < T_1$ 这种限制方法来简化设计, 在性能要求不能满足时再采取一些局部措施。

预充

众所周知, 预充是一种极为有用的电路技术。由两个组合逻辑块和一条预充母线组成的简单系统示于图 6.8, 母线用 ϕ_2 预充, 在 ϕ_1 期间则执行所有母线的输入和输出。在逻辑块内部或从某个外部控制来自通道的逻辑, 要保证在任一周期中只有一个子系统是

1) 实际上是用触发器作为锁存器。

在驱动母线；在 ϕ_1 期间，母线的输入端必须是稳定的，如果在 ϕ_1 为高时，该输入端出现假信号，则母线可能会不可预料地放电。对于图 6.8 上的网络 C 和网络 L，可以看出，在 ϕ_2 期间 C 不可能稳定，在 ϕ_1 期间 L 也如此，另一方面，L 和 C 虽然在 ϕ_2 和 ϕ_1 期间分别具有稳定的可能性，但到此这种可能性也并不是有保证的。因此，这种两相时钟定时的方法需要改变为带有预充的电路。显然，L 网络的不稳定问题是因为其输出端不直接连在预充母线上，必须用一相时钟使输出延迟到能被安全地使用为止。对于 C 网络的输入也可以用同样的方法处理。

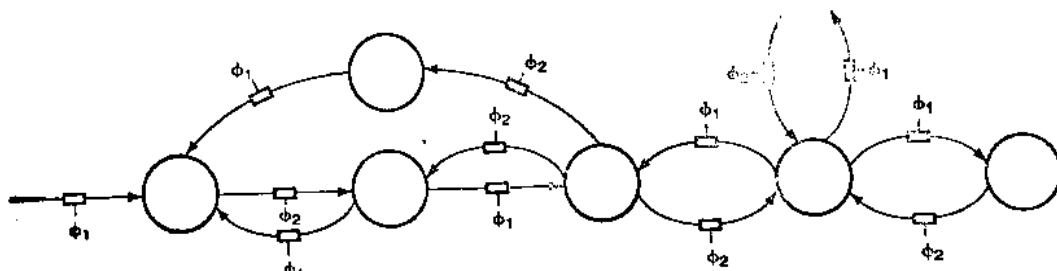


图 6.7 两相子系统。输入或输出都用 ϕ_1 定时

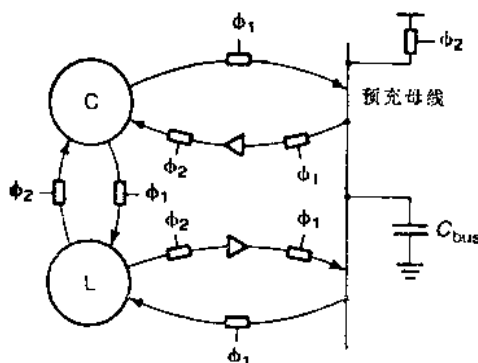


图 6.8 与预充母线的连线。必须阻滞某些信号，同时所有对母线的输入端不能有假信号干扰

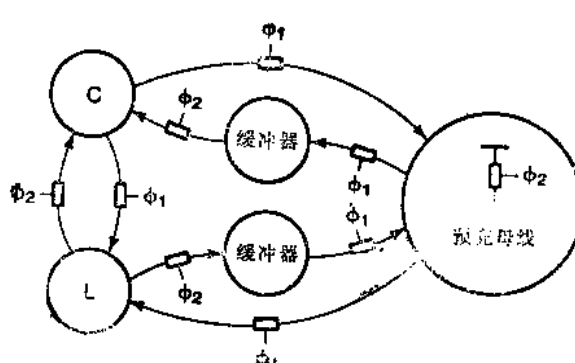


图 6.9 图 6.8 预充母线电路的网络模型

把图 6.8 重新画成类似于图 6.7 的网络形式，结果示于图 6.9。由图可见，除了预充母线外，其余的所有组合逻辑块都用某一相时钟输入数据，同时用另一相时钟输出数据；预充母线则都用同一相时钟 ϕ_1 控制输入和输出。总之，多相时钟电路中有两种不同的模块，对于简单模块，输入信号由 ϕ_i 时钟控制，而输出信号用 ϕ_j 控制，其 $i \neq j$ ；对预充模块，则输入和输出都用同一相时钟控制。信号通过预充模块的延迟必须小于通过简单模块的延迟约束，这样，就找到了一种折衷方案，可以在严格遵守时序约束的条件下利用预充使模块工作更快一些。就是说，预充模块是一种用于两相时钟的电路，而不是用于单相时钟的。由于要求这两个信号互相不干扰，对时序的要求就比每周周期只调用一次的模块更苛刻。图 6.10 所示电路为另一种模块，对它的调用每周周期就不止一次了。

预充模块和简单模块间的相互作用可以用图 6.11 所示的电路来阐明。信号通过预充模块的延迟 τ_c 的约束条件是

$$\tau_c < T_1 \quad (6.8)$$

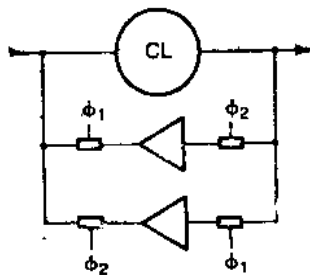


图6.10 一种每周期间用两次逻辑模块的电路

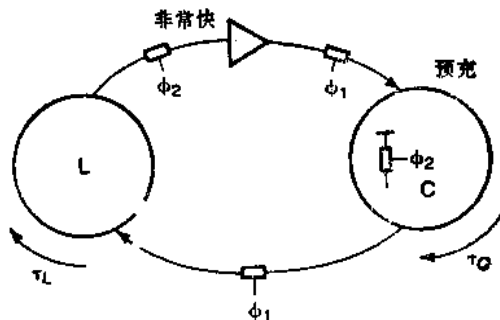


图6.11 带有一个预充模块的电路，该模块的输入和输出都由φ1定时

另外两个约束条件为式(6.6)和式(6.7)，它们构成的约束空间示于图6.12。如果把两个预充模块连接在一起，就得到图6.13的电路，这时需增加另一个约束条件：

$$\tau_L < T_3 \quad (6.9)$$

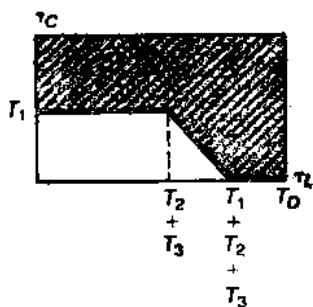


图6.12 图6.11电路的约束空间

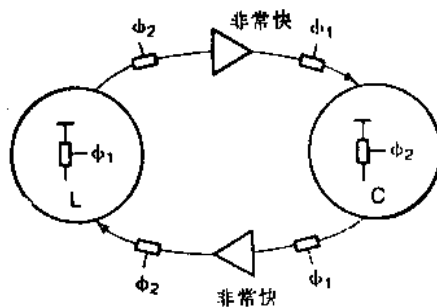


图6.13 一种具有两个预充模块的电路

同时约束条件式(6.5)到式(6.7)在这里就不起作用了。需要注意的是，用了预充模块后自由度减少，图6.14为相应的约束空间。另外，延迟时间还应遵守

$$\tau_L + \tau_C < T_p - (T_2 + T_4) \quad (6.10)$$

其中 $(T_2 + T_4)$ 为无用的时间。尽管时序约束更严格了，但由于我们需要用预充来使大部分nMOS系统的功耗降低，所以在设计时钟波形时总是尽量把非重叠时间 $(T_2$ 和 $T_4)$ 减至最小。

时钟逻辑综合运用了预充技术和移位寄存器技术，图6.15即为一组预充逻辑门，在该系统中，对每一个时钟节拍信号只通过一级逻辑，因此是一种以流水线方式工作的电路。通常采用四相时钟，第一节拍使电路预充，第二节拍进行取值，第三节拍对下一级保持稳定，第四节拍则使信号得以进位一逻辑级，从而同时把信号及其补码送到相同逻辑级的输入端。LSI-11即为这种老式设计的例子。在nMOS还不能生产的时期，就用这种方法来降低pMOS电路的功耗。电荷

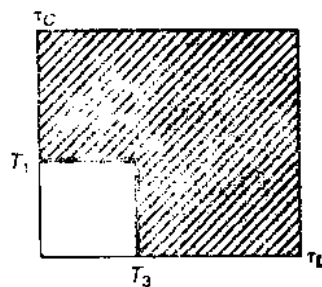


图6.14 图6.13电路的约束空间。由于预充引起的附加约束条件，约束空间已缩小了

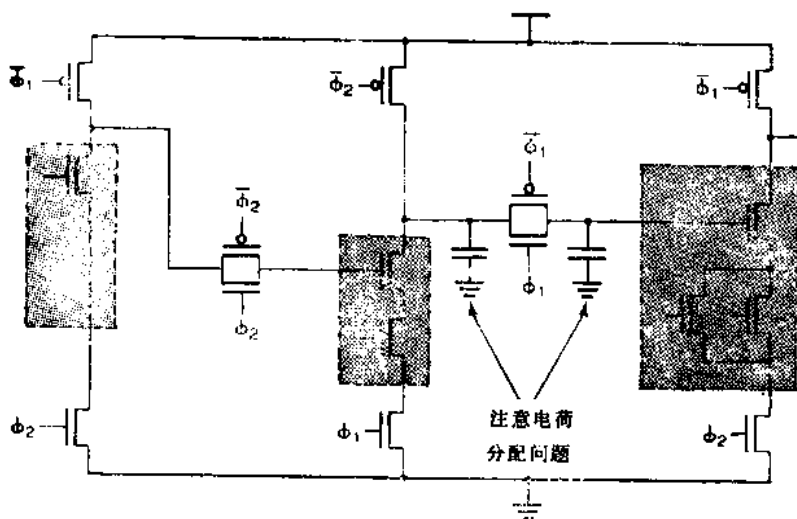


图6.15 流水线式预充电电路，它代表了一种老式设计方案

分配是其设计的关键所在。

例6.1 设计用于图 6.13 电路的 nMOS 缓冲器。

这是一个说明什么情况下可以使用时钟缓冲器的理想例题。例如，我们可以采用图 5.12(d) 中的缓冲器。

把图 6.13 的电路重新绘制，示于图6.16。

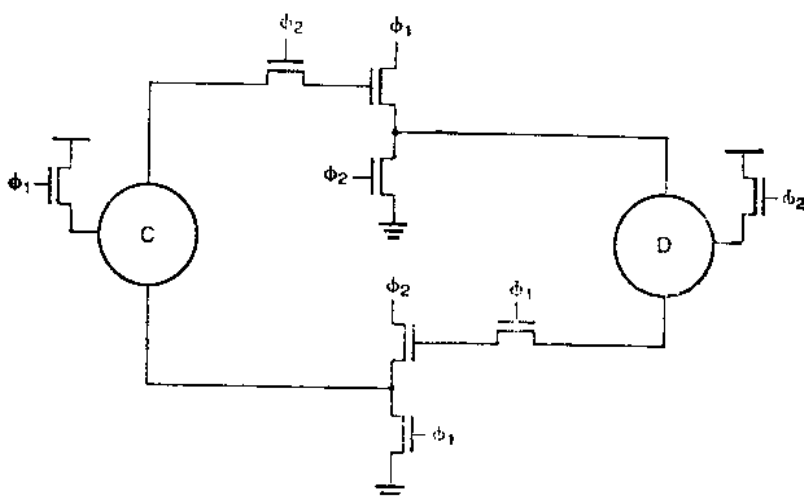


图6.16 用nMOS实现图6.13中的电路

多米诺和行波逻辑

上面分析的预充电电路在每一周期只对一级逻辑预充和取值，当然也可以把逻辑门预充成在每一节拍对多个逻辑级取值。做到这点最直接的方法是仅对第一级逻辑门预充而其余各级则采用静态门。不过，我们也可以采用更好的方案。

若某组逻辑其输入线上电平的单次跳变造成输出端电平也只有一次跳变，那么，我们就可以把这些门组合在一起形成更大的逻辑块。在没有循环的条件下，这些逻辑块也

会具有同样的特性，而这种特性正是预充电路需要的。对逻辑门的预充，只需考虑一次电平跳变就够了。图6.17(a)为预充与非门，当把所有输入端预充为低电平时，输出端则预充为高电平，两者间是存在对应关系的。因此，倘若每个输入端电平出现不多于一次跳变（例如由低到高），则输出端的跳变也只有一次（即由高到低）；这对图6.17(b)的或非门也是同样的。事实上，当n沟的下拉网络是由增强型MOS管并联/串联组成时，任何预充逻辑门都具有这种性质。如果这种逻辑门后面跟随一反相器，则反相器的输出电平也只会有一次跳变（由低到高）。所以，带有反相器的这两种预充复合门之输出的信号正是我们下一个逻辑级的输入所需要的。

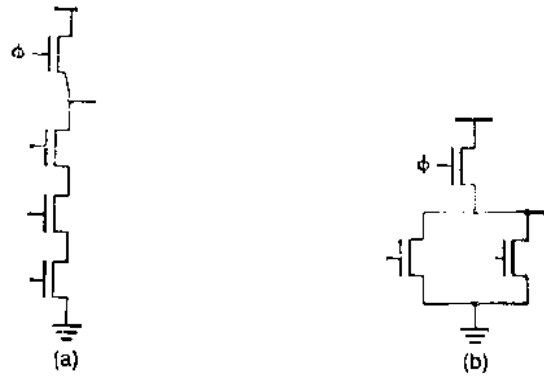


图6.17 两种预充逻辑门。(a)与非；(b)或非

多米诺逻辑[Krambeck 82]是由或门、与门和复合非反相门组成的一种相容的（虽然不是包罗万象的）逻辑系列，由于它的输出是顺序下降，如同一列多米诺骨牌，故得此名。因为每个门的电平跳变不多于一次，电路的动态功耗就降至最低。图6.18为多米诺逻辑电路，图1.70所示的PLA也是多米诺逻辑系列之一。

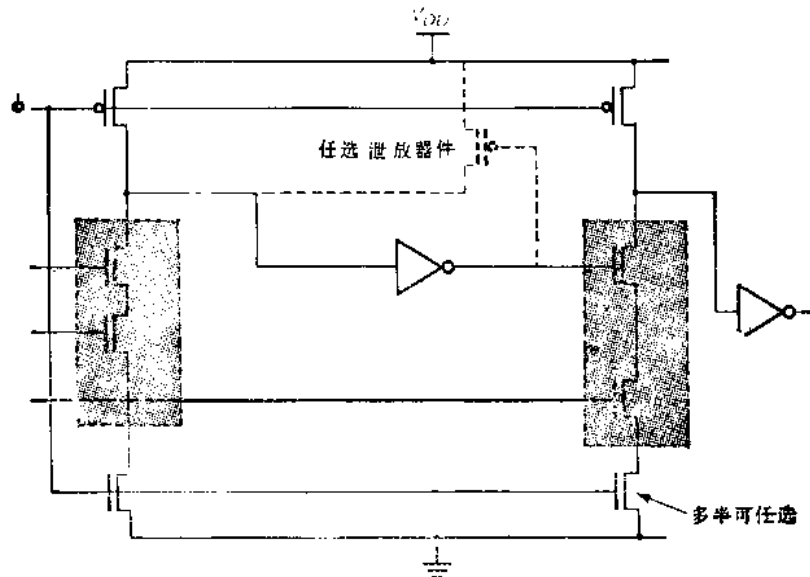


图6.18 CMOS 多米诺电路

传输结构中也可以出现行波。我们将在第八章中分析称之为“曼彻斯特进位链”的技术，这种预充型的电路也呈现一种行波特性。

用对偶原理可以知道，如果把全部 P 型器件换成 n 型器件，同时把所有高电平改为低电平，则可以得到一种基本上是 pMOS 的多米诺逻辑门。此外，我们可以交替使用预充 n 型门和预充 P 型门，如图 6.19 所示[Goncalves 83]。

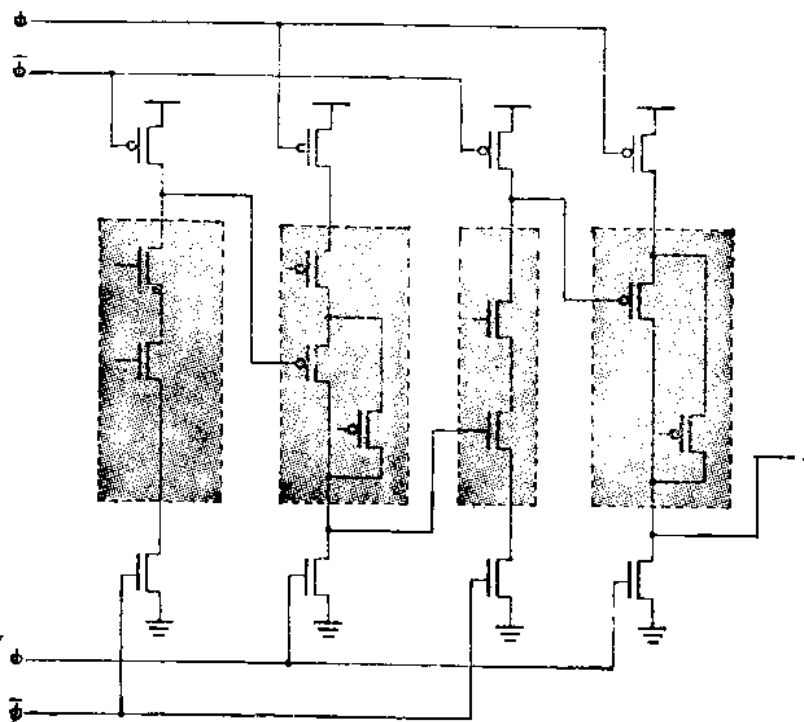


图6.19 NORA 行波逻辑

多米诺逻辑也存在电荷分配问题。电路内部和“与非”图腾柱连接的节点，由于输入为低电平而使它们断开，就可能预充不到高电平，于是当下拉网络中最顶上的晶体管接通时就会出现电荷分配。然而，这问题并不太严重，因为逻辑门的版图设计一般保证了输出节点的电容比内部节点的大得多。此外，栅上电平上升的瞬间引起泵入堆栈的电荷也有助于缓和这个问题。对于十分长的堆栈，内部节点是可以得到预充的。

噪声对图6.19所示的第二类行波逻辑来说是个严重问题。耐噪声强度很低的高阻抗预充节点使这些门对电容耦合噪声极为敏感，特别是当由于性能要求使阈值设置为低值，或因漏感应使势垒降低时更是如此（参考第四章有关噪声的讨论）。

例6.2 重新设计图 6.8 的电路使 C 和 D 都能预充。

解决这个问题需要增加更多的延迟寄存器，如图 6.20 所示。该电路的巧妙之处在于其中有两个部分是在同一时钟相位预充的，只是在有多米诺或行波逻辑电路的条件下才可以把它们直接连在一起，否则，整体时序就需要改变。

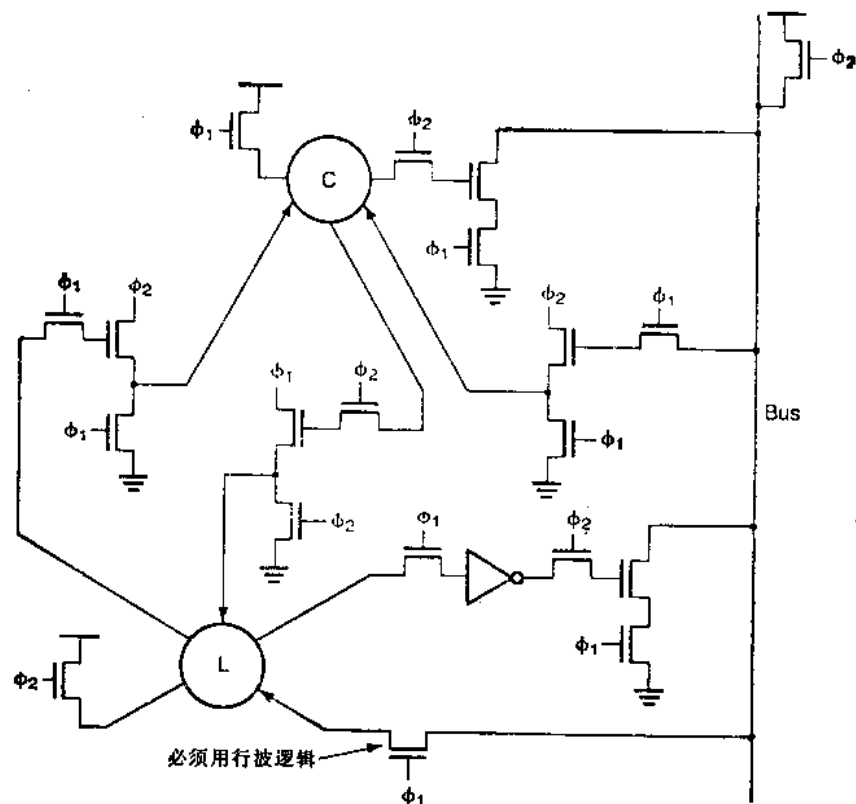


图6.20 用扩大预充的办法实现图6.8的预充母线电路

动态电路考虑

利用传输器件作寄存器显然有一个隐含的条件, 这就是因为电荷存贮机构是动态的, 时钟就不能长时间地停止。另一方面, 为了排除故障, 最好能设计一些单步工作的机器。几种用来组成静态锁存器的电路示于图6.21。

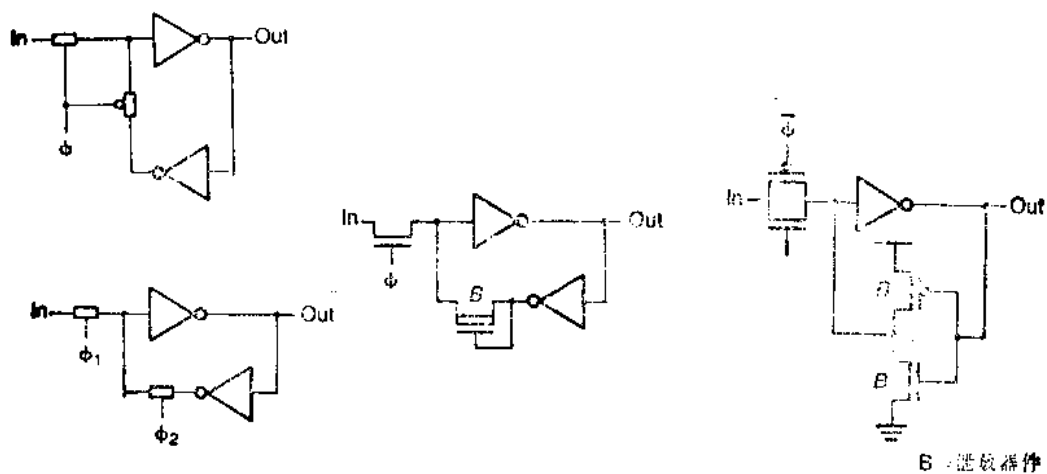


图6.21 静态和半静态锁存器, 在本图中“B”代表二极管

这些电路还可以用于另一些地方。在时钟受到一个逻辑信号控制的情况下，就会有一些动态节点难得被刷新，这种情况示于图 6.22。图 6.23 也属于这类错误，这种电路中当 A 和 ϕ 不同时为真时，电荷就会泵入或泵出动态节点。这两种误操作可以用每周期对所有动态节点进行刷新的时钟定时方法，或者用静态锁存的方法来避免，这后一种方法实际上是不停地对节点进行刷新。

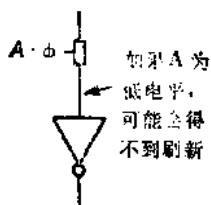


图 6.22 反相器输入端可能会很长时间没有得到刷新

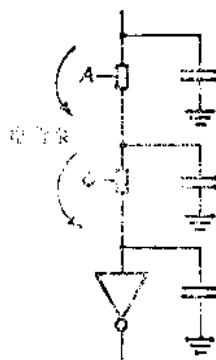


图 6.23 电荷会偶然地泵入或泵出反相器的输入端

从芯片面积和时钟复杂性方面来看，在 CMOS 电路中使用动态传输门比在 n MOS 电路中昂贵得多，所以和 n MOS 相比，在 CMOS 电路中采用静态寄存器成本增加少一些。采用有比 CMOS 锁存设计是有效地利用静态寄存器的一个关键。一种以交叉耦合互补 CMOS 反相器为积木块的设计方法示于图 6.24，该电路在置位和复位网络永不同时作用的条件下只消耗动态功耗。图 6.25 为这种类型电路的一个例子。置位复位网络常可共用晶体管(Heller 84)，也可用预充方法来改善有比电路的上拉速度。

对阵列预充会引起很大的电流尖峰，这是一种噪声源。可以用参差的预充时钟来限制最大电流，在某些情况下，使预充只到 $V_{DD}/2$ 也是一种有用的方法。

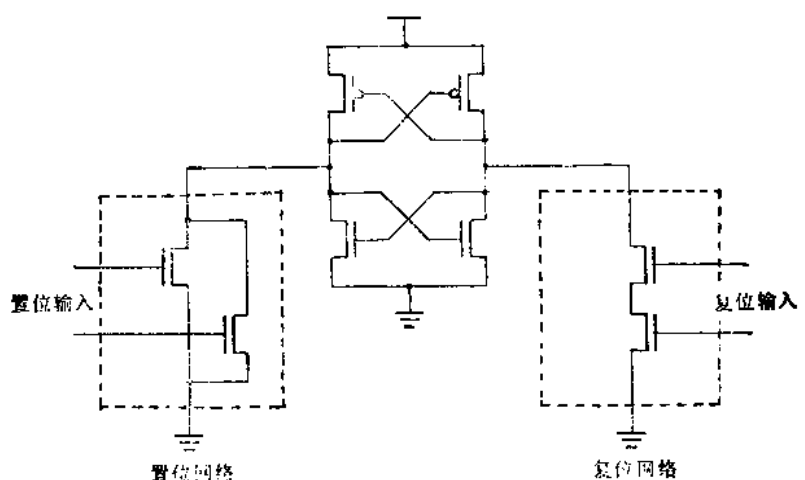


图 6.24 有比CMOS网络，其中置位和复位网络常可共用晶体管

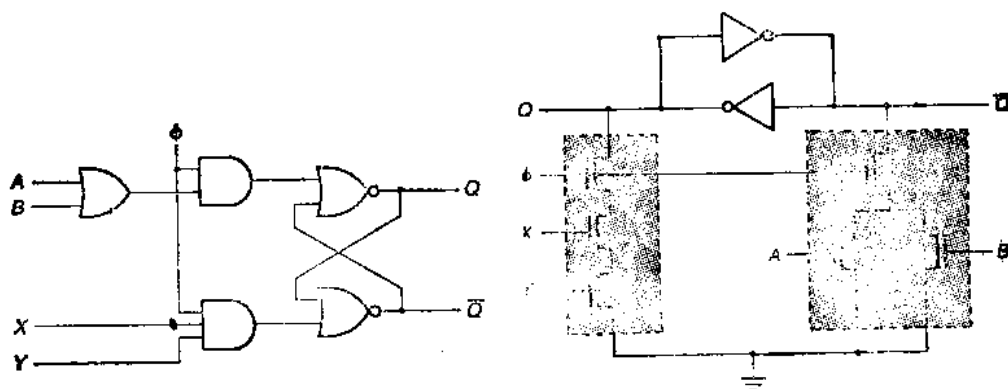


图6.25 用图6.24中的技术实现的某种逻辑。正确操作要求逻辑关系为 $XY(A+B)\phi=0$

时钟歪斜

通常，对大系统来说，控制时钟歪斜是一大难题，而 VLSI 系统则无一例外。实际上，由于 VLSI 系统的时钟跳变时间比一个快速 MOS 门的传输延迟时间长得多，使这一问题更为严重。由于我们不能精确预测时钟的时序，从而降低了 VLSI 系统的性能。在极端情况下甚至会在时钟歪斜问题上增加新的设计约束。

为了分析不可控制的延迟对系统时序约束的影响，考虑一种简单的延迟变换，它并不改变模块总的时序关系。图 6.26 为一个具有任意多个输入和输出的模块，若在每一个输入端增加延迟 d (d 可以是正的或负的)，同时从每一个输出端减去同样数值的延迟，则该模块的时序就保持不变。经过这种变换的模块示于同一图上。

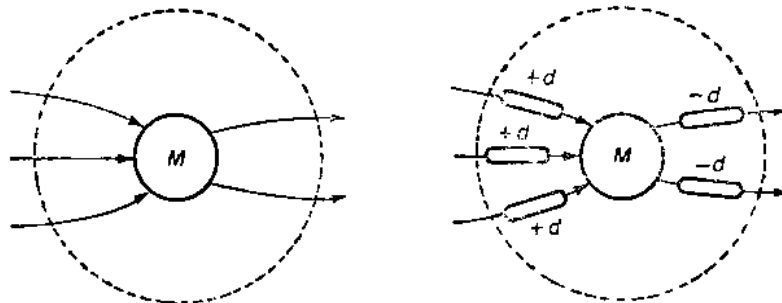


图6.26 在输入端加入延迟，在输出端再去掉该延迟，电路模块就被重新定时了

这种变换使我们有了一种简单的方法来考察时钟歪斜对标准两相非重叠时钟的影响。设某一相时钟有一延迟误差 d ，可以把该延迟误差变换到环路中去，如图 6.27 所示。

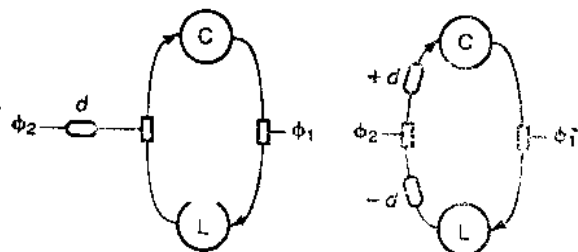


图6.27 对时钟有歪斜的电路重新定时。带有已知通过组合逻辑的延迟而时钟歪斜不稳定的电路，可以把它变换为没有时钟歪斜但通过组合逻辑的延迟为不确定的系统

示。这实际上是把延迟 d 加到一个组合逻辑上，同时从另一个组合逻辑减去同一延迟。根据前面的延迟约束图 6.6，该延迟使 τ_c 轴和 τ_D 轴在 $(-1, 1)$ 方向上位移。然而，我们未曾用过负延迟来表示约束空间，尽管这些负延迟在一个标准系统中是不存在的，但 对一个经过变换的系统来说，它们则确实是有意义的。图 6.28 为两相非重叠时钟系统的全部约束空间，图 6.29 则表示当时钟误差增加时约束空间是如何缩小的。注意在极端情况下时钟就会成为重叠的，从而又回到双边约束条件。

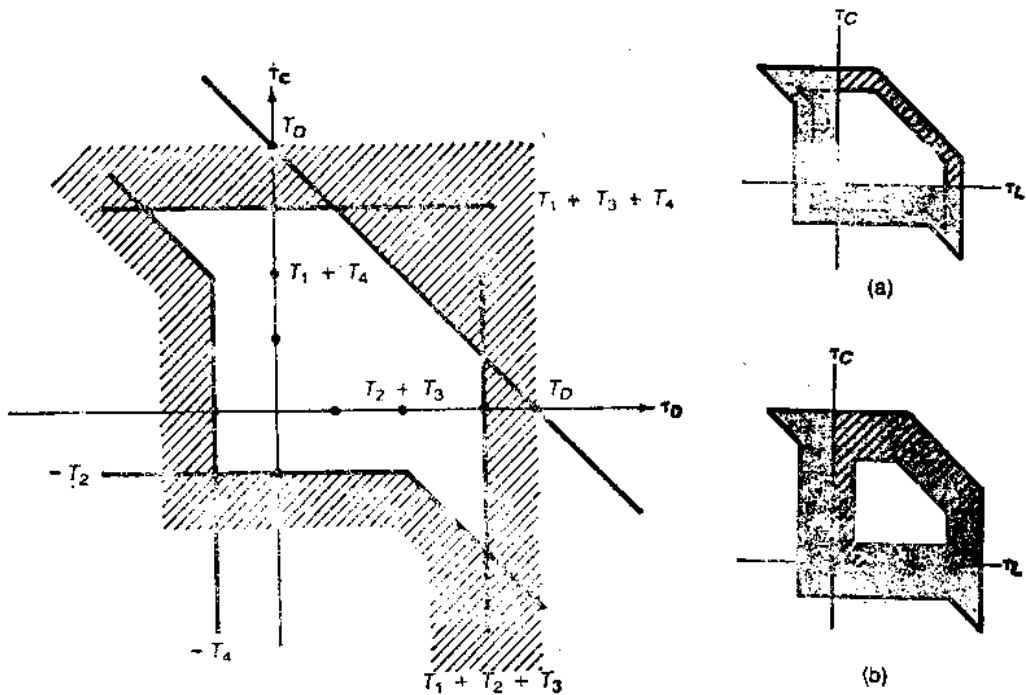


图 6.28 两相非重叠时钟的全约束空间

图 6.29 由于时钟歪斜而使约束空间缩小。(a) 时钟中等歪斜；(b) 时钟歪斜很大，从而导致双边约束

一些与性能和时钟歪斜有关的实际问题出现了。正如我们已经看到的那样，时钟误差会影响模块的最大延迟约束。当工艺参量表现为“慢”时，这种影响也许就明显了。时钟歪斜与逻辑电路的延迟两者之间几乎是没有什么关系的，理解这一点很重要，例如，多晶硅电阻对时钟波形的影响就比对逻辑波形的影响大得多¹⁾，并且，时钟产生电路一般为一些特殊电路，它们的最坏条件区和电路的其余部分是不同的。实际上，在广泛使用预充的条件下，要使时钟之间的无用时间十分短就得付出巨大的额外代价，真正的困难在于保持时钟在“快”工艺条件下不重叠。还需要注意的一点，是在 nMOS 设计中常见的自举时钟缓冲器也具有其本身的最坏条件，这些条件和电路的其余部分也是不同的。

可以用时序容限来计算这一切²⁾，典型的时序容限约为 15%，在电路设计的初始阶段，某些时序容限已包括在对版图寄生参数的保守估计中。

- 1) 在版图设计中，用多晶硅线来分配时钟，往往具有很大的吸引力。
- 2) 如果不显式地采用时序容限，则时序容限就会隐式地超出最坏条件延迟模型，而这些模型是为满足工艺偏差范围而设计的，并且归根结底，其文件也包括了模型误差。

在 CMOS 系统中,除了 ϕ_1 和 ϕ_2 以外还要产生 $\bar{\phi}_1$ 和 $\bar{\phi}_2$, 这是一种额外的负担¹⁾, $\bar{\phi}_1$ 和 $\bar{\phi}_2$ 可以是局部产生,也可以是全局产生,但这两种基本选择没有一般规律可循。全局时钟可以更好地控制引起的时钟歪斜,而局部反相器则连线要求较少。与 nMOS 电路相比,CMOS 电路更呈静态特性,这一点使上述选择进一步复杂了,不过也使我们能更有效地利用时间 T_2 和 T_4 , 此时这两个时间需要更长一些,以确保时钟的“非重叠性”。

例6.3 预充电路。

图6.30的电路中使用了透明锁存器和预充或非门。重画该电路使之和本节中描述的时钟系统类似。图6.31即为这类电路的一种重画形式。

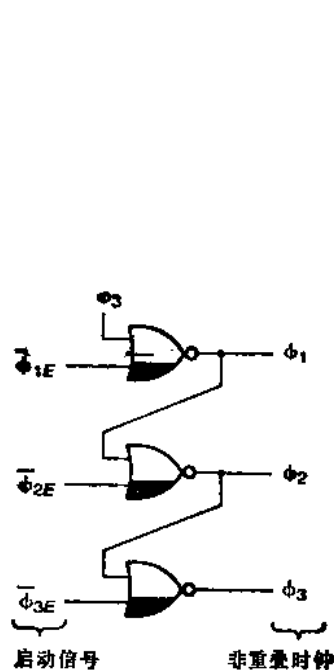


图6.30 一种两相电路实例

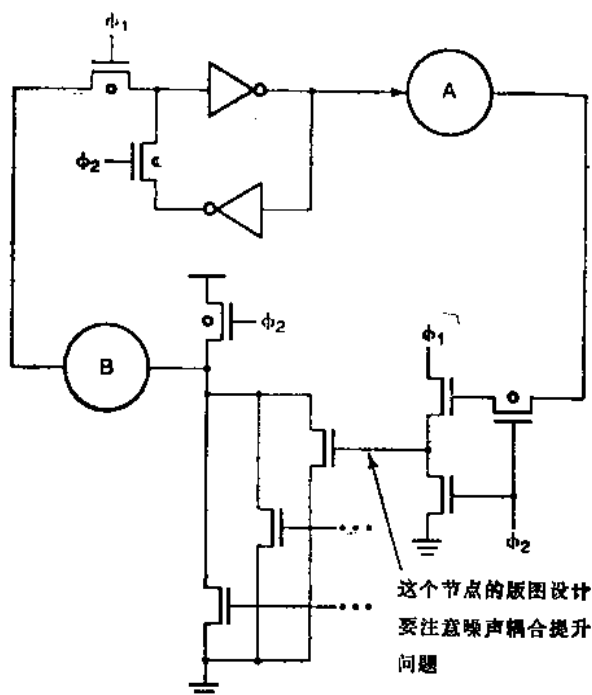


图6.31 图6.30电路的模型

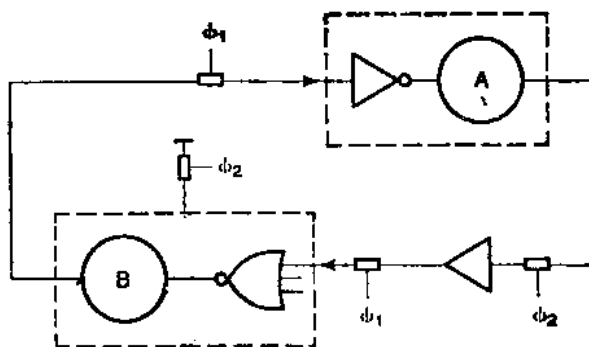


图6.32 图6.30电路的另一例

- 1) 有几种方法试图避免必须产生这些互补时钟, 其一是令 $\phi_1 = \bar{\phi}_2$, 另一是只采用 nMOS 传输器件。虽然这两者都已成功地应用, 但每一种方法都是有严格限制的。

例 6.4 按例题 6.3 的要求重画图 6.32。
结果见图 6.33。

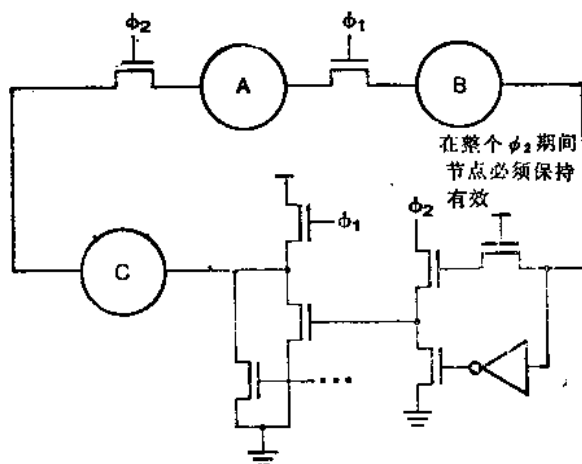


图 6.33 图 6.32 电路的模型

非重叠时钟发生器

我们在 5.2 节中分析了时钟驱动电路，本节中我们要阐明如何把那些时钟驱动电路和逻辑电路组合在一起以构成时钟发生器。

图 6.34 是用于产生非重叠时钟的时钟发生器，反馈保证了时钟的相位不会重叠。在这些发生器中设置了缓冲器后，就可获得一相以上的时钟，例如，图 6.35(a) 就是一个用于带有两相非重叠时钟芯片的时钟驱动器。图 6.35(b) 表明，在 ϕ_1 下降到 $V_{ref} - V_{th}$ 以前，它使晶体管 M_1 保持输出低电平，从而使 C_{boot} 得以充分充电。从启动信号 ϕ_{2s} 变高到 ϕ_1 降至小于 $V_{ref} - V_{th}$ 有一段延迟时间，毫无疑问，这正是明确地保证我们在第五章所说条件所需要的。在 5V 电路中 V_{ref} 可以取 2V 左右。

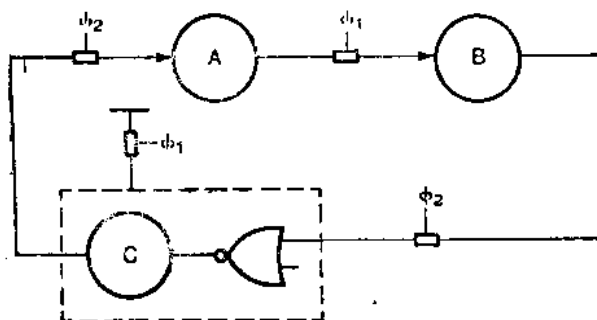


图 6.34 非重叠时钟的一般产生方法

在自举电容充电过程中，为了保持自举节点为低电位，就要求有抗瞬态冲击电流的能力，当输出驱动器不用自举时就不再需要这样做了。由于一般情况下总会有若干个缓冲级，因而可以用选择在那一级发送时钟控制信号来调整非重叠时间总量。图 6.36 为一个三相非重叠时钟发生器的第二相时钟驱动器，这是一个流水线方式工作的电路，线

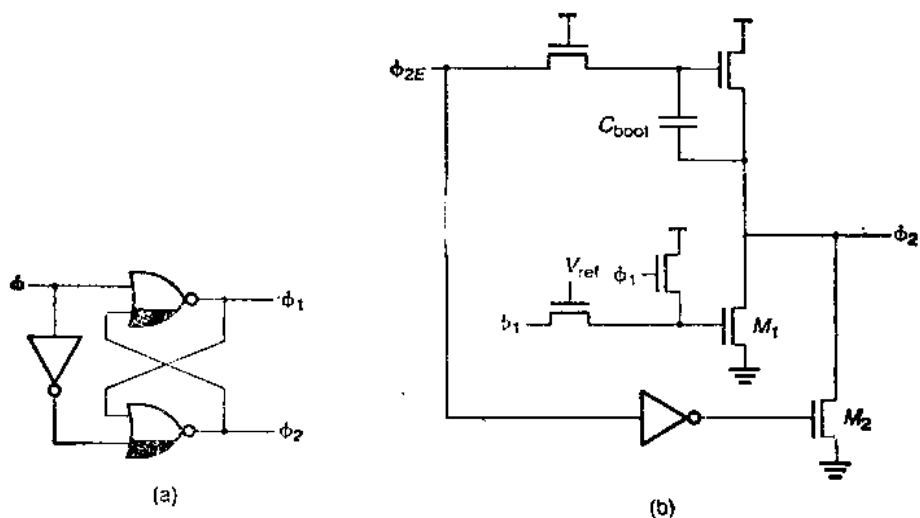


图6.35 (a) 两相非重叠时钟发生器的逻辑图, (b) ϕ_2 时钟驱动器电路。在 ϕ_1 降至很低的 ($V_{ref} - V_{TE}$) 前, ϕ_2 驱动器保持低

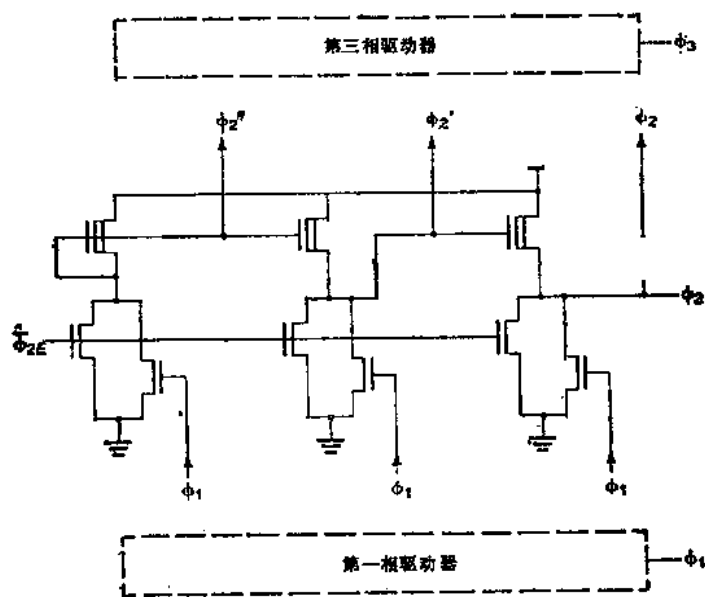


图6.36 流水线式 ϕ_2 驱动器, 它可使非重叠时间减到最小

上的每一级控制第三相时钟驱动器上和它对应的级, 并且受第一相时钟驱动器相应级的控制。

正如我们已在 6.2 节中看到的那样, 两个时钟都为低的时间间隔严重地影响芯片的全部性能, 特别是在采用预充时更为强烈。因此, 维持一个短的, 但是确有保证的非重叠时间间隔就成了产生非重叠时钟的主要困难之一。

6.3 重叠时钟

在所有数字系统中，时钟应当和机器内某些自然发生的逻辑间隔相对应，例如，微处理器的自然周期就是机器的微周期，两相时钟方法则可能把这微周期再分为两个节拍。然而，在许多情况下更为理所当然地把逻辑分成三个或三个以上节拍。

设有一微处理器如图 6.37 所示，数据通道由一个寄存器和一个 ALU 组成，控制通道则为 PLA 驱动的有限状态机。表 6.1 为这种机器常用的时钟定时控制方法和相位分配。

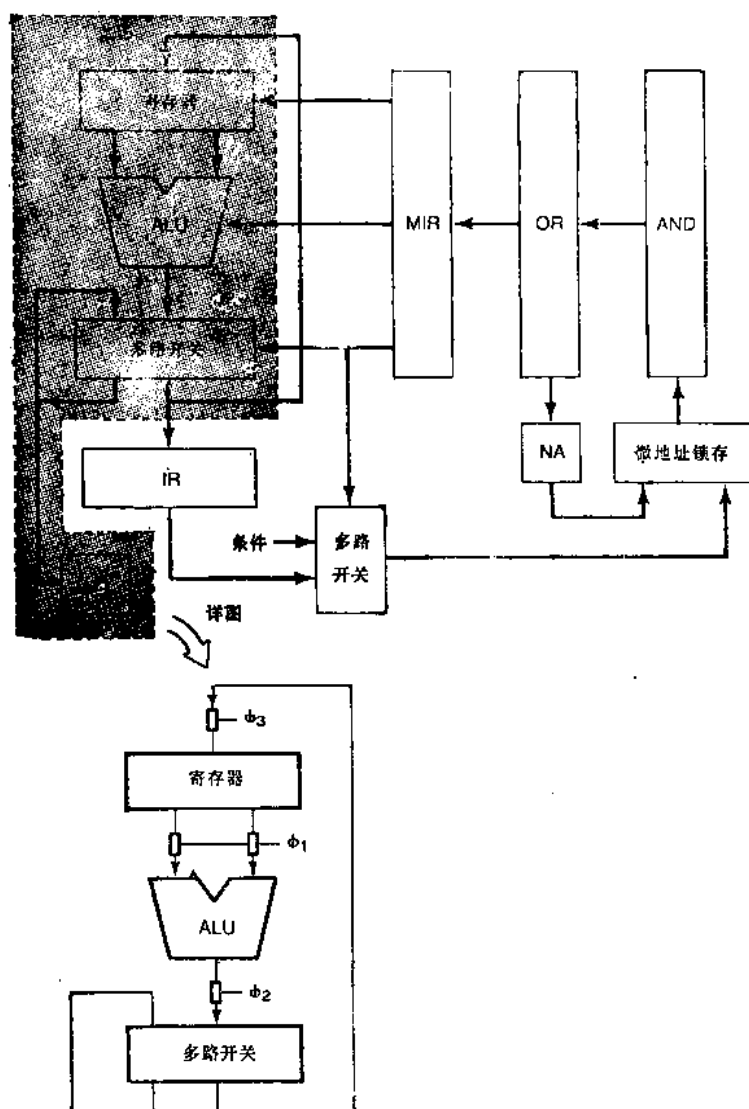


图6.37 用于简单微处理器的三相重叠时钟，在方块图中经常出现时钟控制的传输器件，因为没有别的东西能像传输晶体管那样可以模拟为传输器件

表6.1 图6.37所示微处理器常用的时钟定时方法和相位分配

时钟相位	数据通道	控制通道
ϕ_1	读寄存器	修改微地址锁存
ϕ_2	ALU操作	访问PLA的“与”阵
ϕ_3	写寄存器	访问PLA的“或”阵, 修改MIR

三相时钟法对这种简单的系统是十分合适的。在一个系统中采用三相或更多相时钟的主要优点是避免了产生非重叠时钟, 而这种非重叠时钟在高性能的大系统中是很难设计的。图 6.38 为两相非重叠时钟设计方法的约束空间, 这种方法显然是兼有几个方面的最坏情况。采用三相时钟是最简单易行的重叠时钟定时设计方法。

图 6.39 是一个典型的三相系统及其约束空间。现在约束空间是三维的了, 在全

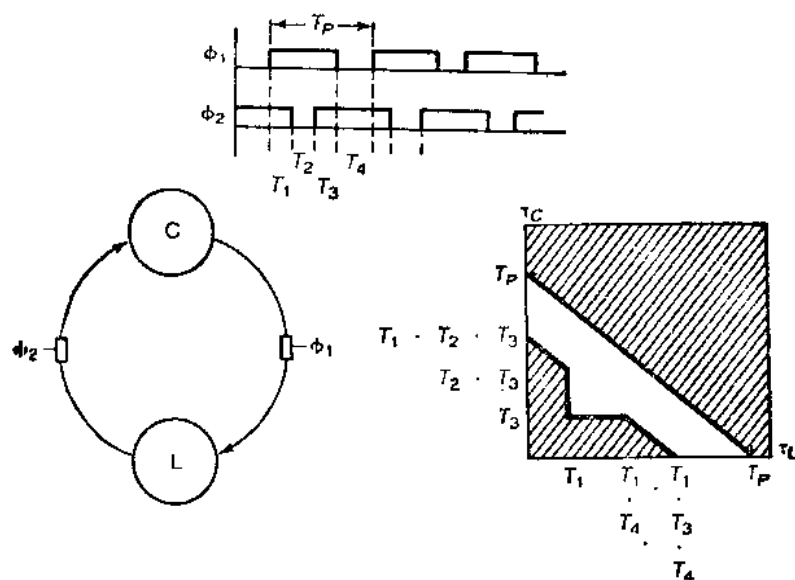


图 6.38 两相非重叠时钟的波形和约束空间。这种方法兼有数种设计技术中的最坏结果

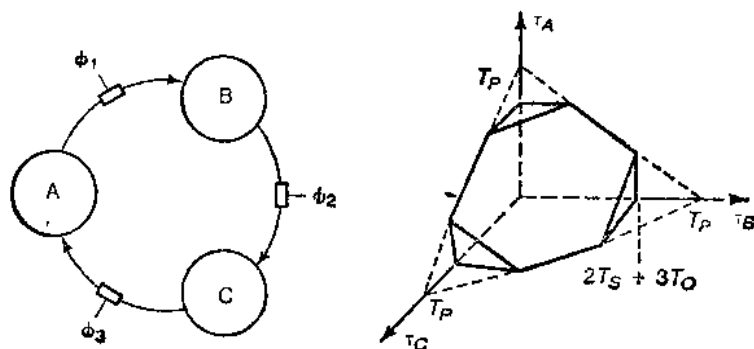


图 6.39 三相重叠时钟的波形和约束空间

三相时钟都不同时重叠的条件下，所有约束都是单边的。即使在时钟有严重歪斜时，这种约束条件也是相当容易保证的（当然，时钟歪斜仍会使性能变坏）。

本节所介绍的 重叠时钟系统，所有周期起码要分为三个节拍。预充的做法与非重叠时钟基本上是一样的，区别仅仅在于存在某一使上拉和下拉晶体管同时导通的 短时间。如前所述，如果在三相系统中用 ϕ_{i-1} 预充一个模块，则它的前面和后面一定有用 ϕ_i 控制的传输晶体管，如图 6.40 所示。

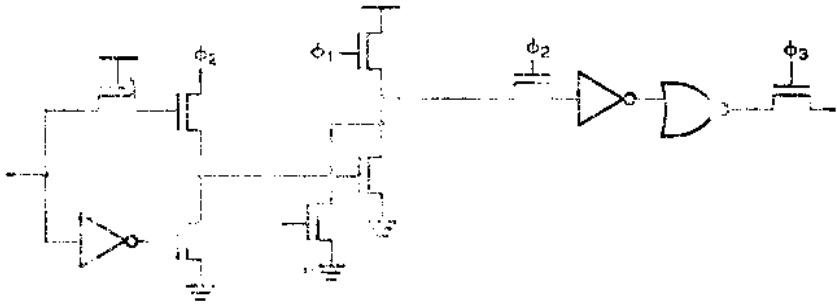


图6.40 预充一个三相系统和预充一个非重叠时钟系统完全一样

重叠时钟发生器

产生重叠时钟比产生非重叠时钟容易得多，这是因为相邻时钟边缘在时序上的不确定性对电路功能的影响不那么重要了。边缘歪斜，特别是预充电路的边缘歪斜，对性能的影响仍是重要的，但是，由于这种重要性仅限于“慢的工艺条件”，约束多少有所减弱。对于非重叠时钟设计方法，时钟是绝不能重叠的，而对一般的重叠时钟设计方法，时钟即使重叠在一起也不会损害电路功能。另一方面，重叠时钟至少需要三相，而对非重叠时钟则至少为两相。

在重叠时钟发生器中，对时钟边缘定时最有效的办法是用 ϕ_{i+1} 的输出中止 ϕ_i 的输出，如图 6.41 所示。虽然从逻辑上看这些输入是可以使输出变低的，但它们实际上没有这种驱动能力。图 6.42 是一个典型的相驱动器。启动信号可以用约翰逊计数器产生，图 6.43 就是这种约翰逊计数器，其中的 D 触发器应该是静态的。

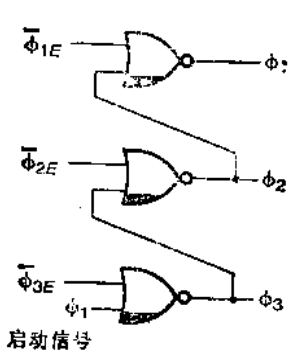


图6.41 三相重叠时钟发生器，每一相时钟中止它前面的那一相时钟

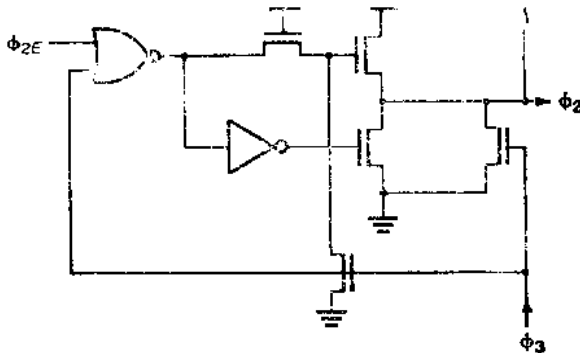


图6.42 ϕ_2 时钟缓冲器的详细结构图

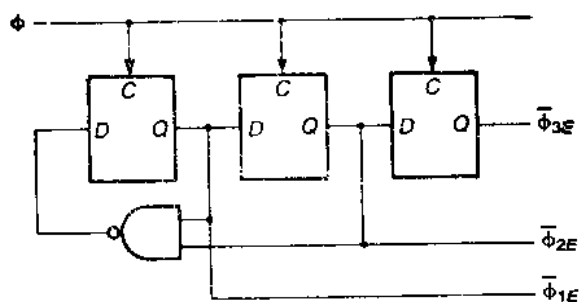


图 6.43 用于产生启动信号的约翰逊计数器，锁存器应为静态的

6.4 自定时 MOS 电路

自定时电路由于其性能良好和元件使用效率高而经常被用于 MOS 逻辑电路。自定时电路控制机构的核心部分是由两个信号和一个状态变量组成。第一信号是 MOS 逻辑功能的一个输入，并启动计算过程（“GO”）；第二信号则是逻辑功能的一个输出，同时标志计算过程完成（“DONE”），这第二信号也可以作为下一逻辑块的“GO”信号。表示逻辑是否在工作的信息则存贮在一个锁存器中。图 6.44 为自定时逻辑功能的方块图。GO-Compute-DONE 周期有时候修改成包括复位功能，即为 GO-Compute-DONE-RESET-GO …。这种完成信号可以用下列两种方法产生：

1. 使用一种和过程完成的检测结合在一起的信号电路。
2. 用一种模拟延迟来仿真要执行的逻辑操作。

对于第一种方法，整个系统一般是异步的，同时为了实现信号电路而附加的逻辑需要额外地增加相当大的版图面积 [Mead 80, Chapt 7]。这种方法的潜在优点是由于“DONE”信号是用与数据有关的方式产生的，所以可以利用某个特定问题的统计性质。然而，LSI 系统很少采用这种方法。

MOS 系统几乎总是选用第二种方法，这是因为它的元件效率高。例如，对于阵列结构，可以用一个每周都工作的虚元件来产生完成信号。这种方案的主要缺点是必须把虚元件设计成比最慢的逻辑通道还要慢一些。当用该完成信号来触发下一级电路时，延迟将大于那些非最坏条件下设计的通道所必需的值；不过，阵列结构的通道之间延迟的差别并不大，这一点是要注意的。

对于第二种自定时逻辑形式，其主要之处在于我们能够用一个小得多的电路来模拟通过一个大的逻辑块的延迟。之所以能这样做，是因为在一块 LSI 芯片上工艺参数是一致的。这些自定时技术常在同步系统中使用，且被认为是用来产生芯片上某一部分所需要的额外时钟的巧妙方法。例如，我们可能会要求在 ϕ_1 期间既进行预充又进行逻辑

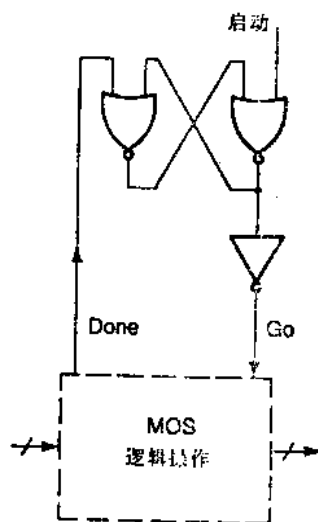


图 6.44 自定时 MOS 逻辑功能

取值,但却没有必需的额外时钟,利用自定时方法,就可以通过检测何时预充完成来产生该时钟。在这种情况下,如果我们从最坏条件来模拟,以保证在 ϕ_3 开始以前完成预充和取值,则系统仍然是同步的。动态RAM大量使用自定时时钟。

例 6.5 自定时数据通道。

在nMOS微处理器数据通道中,有许多地方可以采用自定时逻辑。前面介绍了三相数据通道, ϕ_3 写数据通道之后是 ϕ_1 读。一种双母线标准六管存储单元电路示于图 6.45。

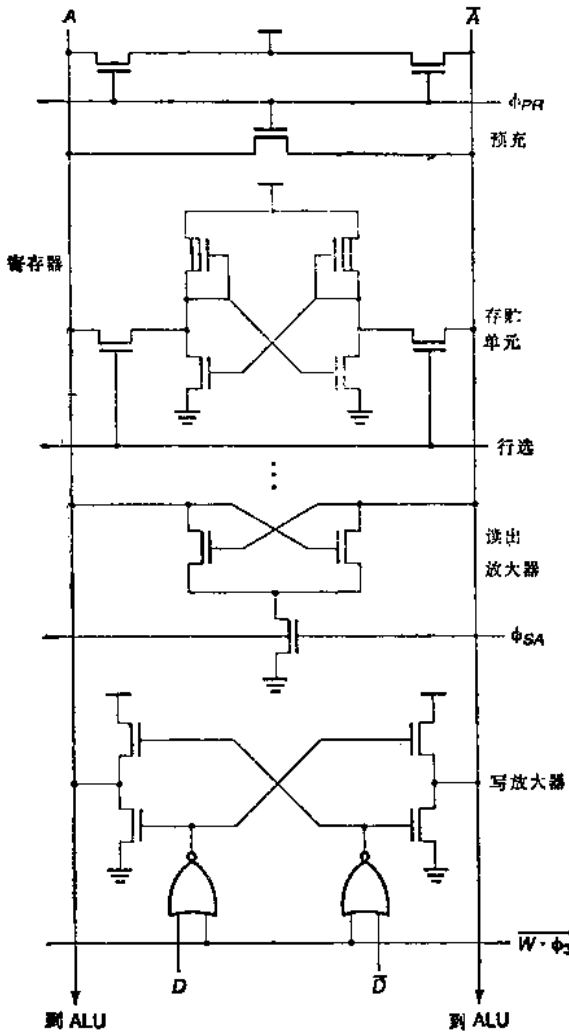


图 6.45 一种必须在 ϕ_1 执行若干操作的寄存器

元去选,从而预充时间又可以用来把新的行选数据送入行驱动器。执行这种任务的逻辑示于图 6.48,而图 6.49 为其时序波形,其中第N行为上一周期选中的行,新周期选中的行则为第N+1行。

余下的问题只是在恰当的时候触发读出放大器了。读出放大器的时钟 ϕ_{SA} 应该只有当行选已被再次选中,以及被选中单元数据线上的电平已拉开到相当的噪声容限之后才能开始上升。达到这个目的的控制方法很多,最简单的方法是增加一个虚拟的行驱动

图 6.45。

写放大器在 ϕ_3 期间以差分方式驱动母线,同时把信号写入 ϕ_3 期间选中的存储单元。当我们希望能在 ϕ_1 开始对某一行进行存取时,正确的操作过程如下:

- 除去以前选的行
- 预充和平衡数据母线
- 选择一新行
- 启动读出放大器

全部这四种操作都能用自定时方式在单个 ϕ_1 期间完成。首先我们必须要有自定时预充电路。假设每个机器周期总有一个单元被选中,则母线一定是处在1:0或0:1状态。预充是否完成可以用一个连在A和 $\bar{A}$ 上的简单与非门来检测,若设所有的列都是相同的,则检测器只需要对一对线进行操作。如果用类似于前述三相重叠时钟方法产生时钟,相应的预充时钟发生器如图 6.46 所示,其时序图则为图 6.47。

其次,我们讨论行选与预充互相作用的问题。前面分析过,在 ϕ_3 期间我们一直在按照某些条件写一个先前选中的单元,并且在 ϕ_1 期间读一个新的单元。在预充期间要使全部单

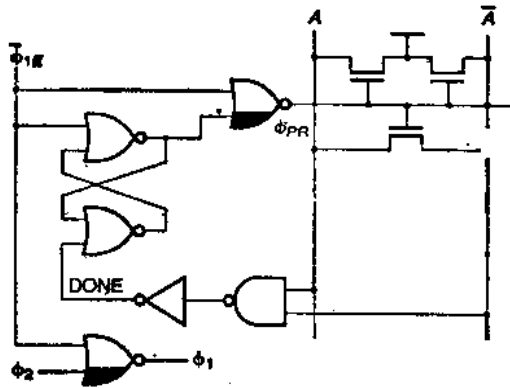


图6.46 自定时预充逻辑

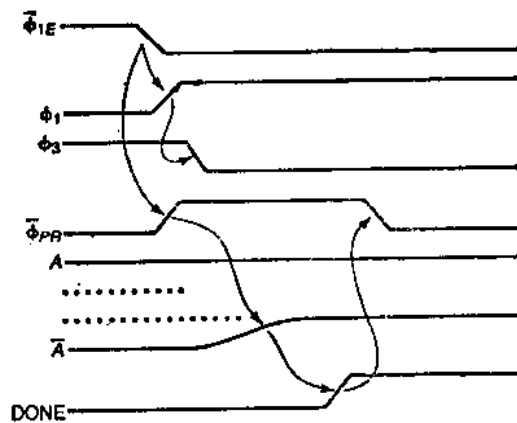


图6.47 预充逻辑的时序

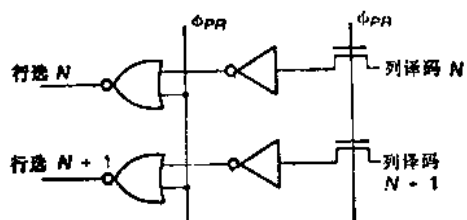


图6.48 行选逻辑

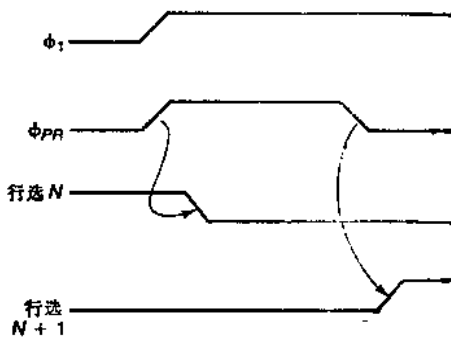


图6.49 行选逻辑的时序

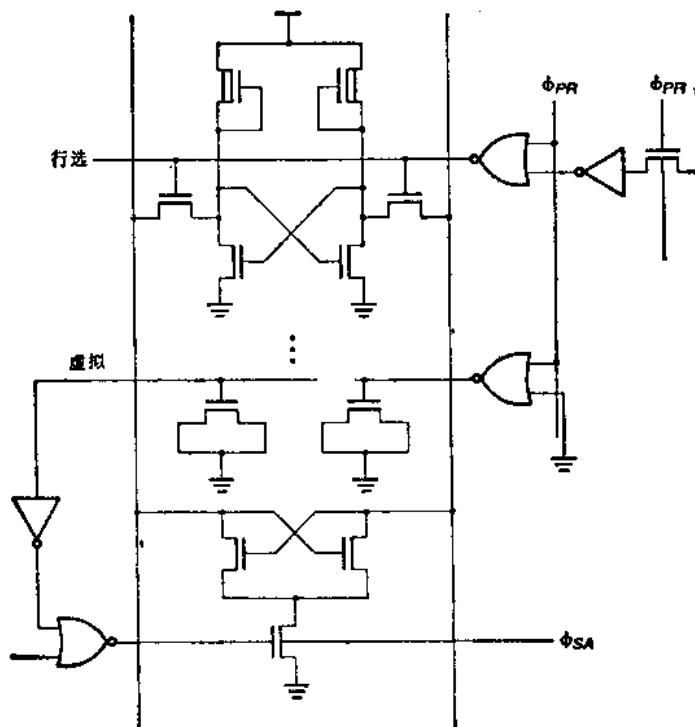


图6.50 触发读出放大器的逻辑

器，同时模拟电容负载或者在阵列中真正放入一完整的虚行。第二种方法示于图6.50。如果采用多晶硅作行连线，则完整的虚拟行是特别合适的方法；在这种情况下，触发读出放大器时必须考虑多晶硅线的延迟。图 6.51 是完整的时序图。

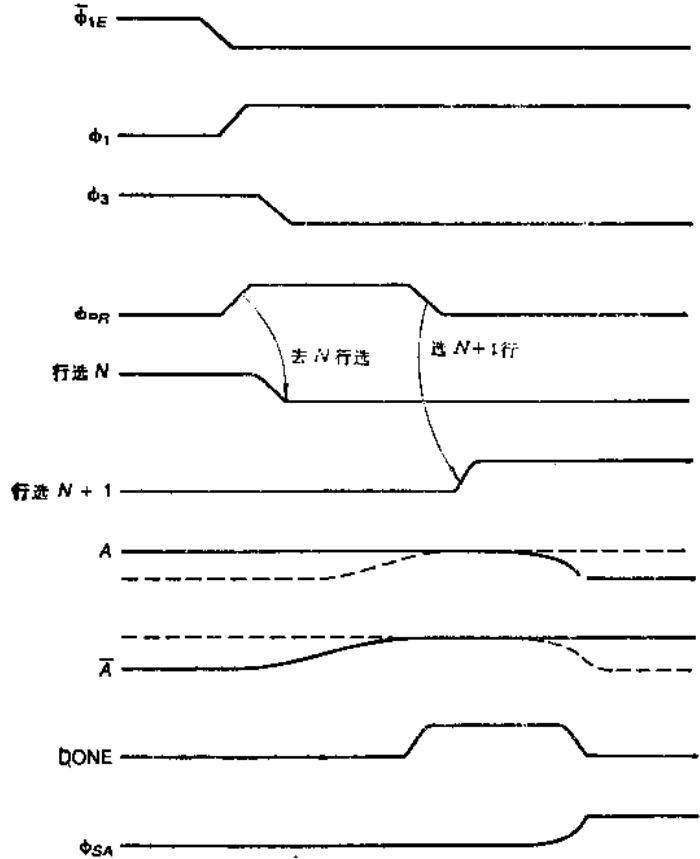


图6.51 自定时寄存器的完整时序

在 ϕ_2 开始之前，可以推测 A 和 $\bar{A}$ 必定是稳定的。但是，设计前面描述的那些电路还是相当棘手的，因为很难证明它们的最坏条件是一致的。事实上，最好用统计模拟法来设计这些电路。我们讨论过的那些自定时电路，其成功更多地与正确的硬件模型（片内延迟线）有关，而不是与软件模型（电路模拟用的模型）相联系，这一点是吸引人的。

6.5 同步器和判断器

大多数同步数字系统最少有一个异步输入（这种异步输入可能在与时钟边缘相关的任何时刻来到），因而必须对这些输入进行同步。举一个两异步事件的实例，一个事件是计算机对某个终端进行访问以查看是否已有键盘输入，另一事件是键盘输入。对于操作员的键盘操作，没有任何理由要和计算机的内部时钟有某种时序关系。我们说计算机极类似于一个同步器，这是指它有一个内部时钟来管理某些必定要发生的事件，所以它是有时间感的。当硬要同步系统去处理非同步信号时，可能会遇到极大的困难。我们再

以计算机处理终端操作为例，若在计算机访问终端时恰好操作员按下某个键，在这种可能性出现时，不管计算机作出肯定（即信息键入很及时，并被记录下来）或否定（即键入不是时候，信息将在下一次访问终端时加以记录）判断，对任何人来说都是无关紧要的。苛刻的问题是计算机必须作出这种或那种判断，而这对计算机来说可能需要任意长的时间。如果计算机尚未作出决定，则在它的内部将会有有一个在高电平和低电平之间的信号出现。当然，该信号本身不会引起困难，问题在于当两个不同的门对该信号作不同的理解时，将引起互相矛盾，并可能导致发生错误（例如把寄存器 $R0$ 读到母线上，去，并和寄存器 $R2$ 读到母线上作比较，就会产生错误，当两者在母线上“与”时就无意义）¹⁾。这种事件的几率虽是很小的，但键按得时间越长，出现的几率就越大。

一般情况下，需要一种电路来对两个事件中的某一个是否已经发生作出判断，这种电路就叫做同步器或判断器 [Couranz 75, Kinniment 76, Kung 82, Veendrick 80]。所以，判断器是一个用来判定两个事件中哪一个首先发生的逻辑电路。同步器是判断器的一个特例，它必须在某个时刻作出判定，所以当判断器的输入之一为时钟时，它就是一个同步器。从传统上说，判断器是用交叉耦合的寄存器来实现的。下面我们分析如何设计一个好的判断器。前面已经看到，判断器判断时是需要一定时间的（例如需要 3 秒钟），因而首先必须明确这个“好”字的定义。由量子力学和测不准原理可知，几乎任何事情都存在一定的发生几率，例如，在本页上的那些字也许会自行重新排列，以至于某些字变为倒过来写了。不过，我们实际上并不担心出现这种情况，因为这种几率小到使事件不象会发生。于是我们探讨的是要使一台计算机因某种几乎不可能发生的、从而我们也用不着担心的异步事件而导致故障的问题。这件事做起来简单，但却昂贵。我们将看到，一个异步事件引起的出错几率，是随时间呈指数下降的。为了使系统工作可靠，我们必须等待很长时间，对于本书所举的 $2\mu\text{m}$ 工艺的例子，这时间长达数十个纳秒 (ns)。这样，对于一个两相或三相系统，常常必须取一个节拍来判别一个异步事件。这对我们讨论的终端应用显然没有什么关系，但对例如两个高速协处理器之间的数据通信来说则是十分重要的。正因为如此，只要有可能，大多数现代计算机系统的母线都做成同步的。

下面我们研究两种不同的情况，首先分析简单的无噪声同步器电路，然后研究噪声对同步器性能的影响。

同步器电路

图 6.52 是一个简单的判断器电路，其关键是交叉耦合锁存器。把某一信号送入锁存器，当 ϕ_1 变低时对输入信号进行采样，该信号可能是高电平，也可能是低电平，或许是介于高电平和低电平之间的某个电平。如果输入信号的电平高于锁存器的亚稳点，锁存器将最终倒向和输入信号是实实在在的高电平时的状态一致的状态。反之，当输入电压低于亚稳点的电压时，锁存器状态与输入低电平时一致。假定输入电压以变化速率为 dv_{in}/dt 通过亚稳点附近区域，且 $v_{in}(t)$ 与 $\phi_1(t)$ 没有时序关系，则 $v_{in}(t)$ 可以是任意值。输入电压在亚稳点 V_{MS} 附近展开，示于图 6.53，图上也标出了 V_{TH} 和 V_{IL}

1) 举一个经典的判断例子。两个站在门厅中间的人相互朝对方行走，当他们企图避开对方的路线时就会出现复杂的（费时间的）摇摆现象。这就体现了一种判断问题。

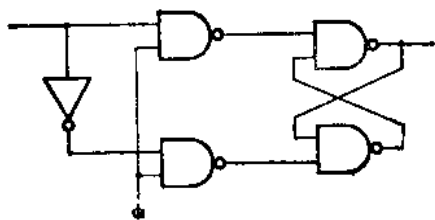


图6.52 同步器

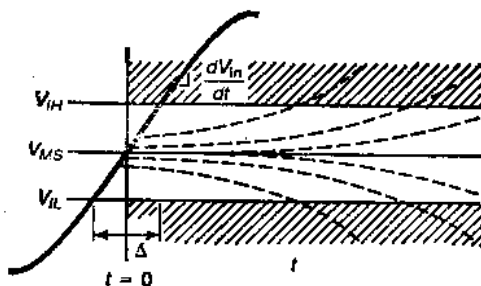


图6.53 同步器波形图，它表示输入电压被锁存后输出电压的可能轨迹

两个电压。输入电压由 V_{IL} 过渡到 V_{IH} 所需的时间 Δ 为：

$$\Delta = (V_{IH} - V_{IL}) \left(\frac{dv_{in}}{dt} \right)^{-1} \quad (6.11)$$

如果等待时间为 T ，那么发生错误的几率有多少？首先分析 $T=0$ 的情况，这时当输入信号是 V_{IH} 和 V_{IL} 之间的某个值时就会出现错误。令同步条件下电平的变化速率为 f_{in} ，因而事件之间的平均时间为 f_{in}^{-1} ，这就是说，若我们不作任何等待，则任一得不到正确地同步的事件的几率 p_E 为：

$$p_E = f_{in} \Delta \quad (6.12)$$

由于我们已假定了该几率是很小的，一秒种内产生错误的几率 p_s 就为：

$$p_s = f_s p_E = f_s f_{in} \Delta \quad (6.13)$$

其中 f_s 为采样频率，所以每秒钟有 f_s 个出错的机会。这样，出错的速率为 p_s ，平均失效时间就为 p_s^{-1} ，或

$$MTF = \frac{1}{f_s f_{in} \Delta} \quad (6.14)$$

以上分析对 $T=0$ 有效¹⁾。当 $T>0$ ，错误率就大大降低。

同步器（设为单主极点）线性区的电压遵从方程

$$v(T) - V_{MS} = (v_{in}(0) - V_{MS}) e^{-T/\tau_{switch}} \quad (6.15)$$

其中 τ_{switch} 为系统的时间常数， $v_{in}(0)$ 是 ϕ 变低时的初始条件， $v(t)$ 则为寄存器电压随时间的变化。

可以认为，若当 $T=t$ 时电压在 V_{IL} 和 V_{IH} 之间，则在 $t=0$ 时，该电压必然处于下面两值之间：

$$V_{MS} + (V_{IH} - V_{MS}) e^{-T/\tau_{switch}} \quad (6.16)$$

及

$$V_{MS} - (V_{MS} - V_{IL}) e^{-T/\tau_{switch}} \quad (6.17)$$

式 (6.16) 和式 (6.17) 之差为一新的 Δ 有效值，于是出错率 $p_s(T)$ 就变为：

$$p_s(T) = f_s f_{in} \Delta e^{-T/\tau_{switch}} \quad (6.18)$$

相应的 MTF 为：

$$MTF = \frac{e^{T/\tau_{switch}}}{f_s f_{in} \Delta} \quad (6.19)$$

1) 到此为止，我们忽略了通过锁存器的传输时间。

可以看出, 系统的可靠性是随时间而指数性增加, 已知要求的 MTF 后就可以求出 T 值,

$$T = \tau_{\text{switch}} \ln(\text{MTF} f_s f_{\text{in}} \Delta) \quad (6.20)$$

我们举一例子来说明, 若要求 $\text{MTF}=10$ 年, $f_{\text{in}}=1\text{MHz}$, $f_s=10\text{MHz}$, $\Delta=10\text{ns}$, $\tau_{\text{switch}}=1\text{ns}$, 由于 $10\text{年}=3 \times 10^8\text{s}$, 得 $T=31\text{ns}$, 它约是 30 个时间常数。

确定 f_s , f_{in} 和 Δ 的值是相当直接了当的, 但如何得到 τ_{switch} 值呢? τ_{switch} 是电路增益带宽乘积的函数, 可以用解析方法求得 (参看 5.4 节)。另一种求 τ_{switch} 的方法是利用模拟, 这种方法也许更为精确。如果我们选取模拟初始条件是使锁存电压十分接近于亚稳点, 则衰减时间常数正好就是 τ_{switch} , 因而有

$$\tau_{\text{switch}} = (t_2 - t_1) \ln^{-1} \left[\frac{v(t_2) - V_{MS}}{v(t_1) - V_{MS}} \right] \quad (6.21)$$

T 值是用这样求得的 τ_{switch} 值计算得到的。需要注意的是, 由于 MTF 和 τ_{switch} 呈指数关系, τ_{switch} 的误差就十分重要了。所以要格外小心处理包含在 T 中的误差容限。 T 也应包括通过判断器电路的延迟, 该值一般为几个 τ_{switch} 。

同步器的故障是极难跟踪的, 因此, 系统中同步器的数量应尽量少 (即不用或用一个同步器)。

为了避免同步器问题, 有时候可以使系统锁相 [Gardner 79], 使系统不再具有不确定的时序关系。锁相技术经常在一些大系统中使用, 在这些系统中, 容易对精确的频率信息而不是相位信息进行分配。当两个系统具有互不相关的然而却是完全确定的频率时, 可以用预测将要发生的同步事件并提前予以解决的办法来避免许多同步器的时序损失 [Ward 83]。换句话说, 一种不可预测事件发生的速度是与其带宽而不是与其频率有关。同步器是用来消除那些不可预测的事件, 所以, 它的性能基本上不是受输入信号的频率而是受这些信号的带宽限制的。

噪声对同步器性能的影响

低幅度的噪声并不影响判断器电路的性能。由于这和直观结论有矛盾, 我们将说明为什么是这样的。这就需要证明任何一种噪声加入到系统中去后并不改变本节中导出的公式, 即要证明噪声是一种完全正交的问题。

在亚稳点附近使电压按对数尺度分割, 例如分成 1 至 $10\mu\text{V}$, 10 至 $100\mu\text{V}$ 等等, 如表 6.2 所示。

已知初始电压可以取任一值, 由于 10 至 $100\mu\text{V}$ 的电压区间比 1 至 $10\mu\text{V}$ 的电压区间大了 10 倍, 当 p_0 为电压处于 1 至 $10\mu\text{V}$ 区间的几率时, 处于 10 至 $100\mu\text{V}$ 间的几率

表 6.2 初始的几率分布

电 压 范 围	几 率
1 至 $10\mu\text{V}$	p_0
10 至 $100\mu\text{V}$	$10p_0$
100 至 $1000\mu\text{V}$	$100p_0$
1000 至 $10000\mu\text{V}$	$1000p_0$

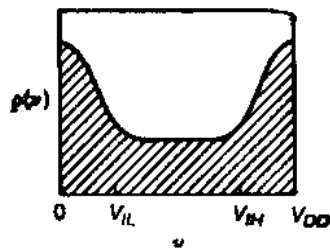


图 6.54 初始几率分布

就为 $10p_0$ 。图 6.54 是初始电压的分布几率。这种几率随时间如何变化是一个重要的问题。系统电压的变化要服从方程 (6.15)，因此，电压从 $1\mu\text{V}$ 上升到 $10\mu\text{V}$ 造成的偏移量和从 $10\mu\text{V}$ 提高到 $100\mu\text{V}$ 的偏移量完全相同，并可以此类推。所以，根据表 6.2 给出的初始几率，在 $T = \tau_{\text{switch}} \ln 10$ 时的几率则如表 6.3 所示。

表 6.3 $T = \tau_{\text{switch}} \ln 10$ 时的几率分布

电 压 范 围	几 率
1 到 $10\mu\text{V}$	$0.1p_0$
10 到 $100\mu\text{V}$	p_0
100 到 $1000\mu\text{V}$	$10p_0$
1000 到 $10000\mu\text{V}$	$100p_0$

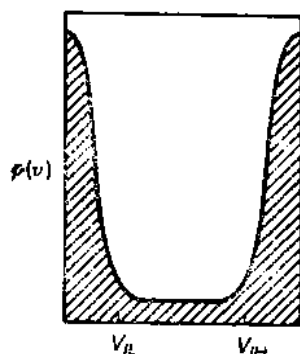


图 6.55 $T = \tau_{\text{switch}} \ln 10$ 时的几率分布

我们已经指出，在 V_{MS} 附近的线性区中，几率密度是均匀下降的，图 6.55 就全面地表示了这种情况。如果初始电压处于 V_{MS} 和处于 $V_{MS} + 1\text{mV}$ 的几率是差不多的，则在两秒钟以后情况依然如此，虽然这时两点的电压都可能会比两秒钟以前小几个数量级。

这一点是很重要的，因为如果在采样时电压处于 V_{MS} 附近的某个值，则只要它是在线性区，对任何 $T > 0$ ，电压仍为该值。分析了这些后我们就能理解噪声的影响了。小幅度噪声的作用相当于使偏置偏离，但是由于系统是在完全相同的情况之间移动，也就无所谓是有害还是有益的了。例如，噪声可能使亚稳电压增加 1mV ，我们已经证明了这并不产生任何变化。

作为为什么噪声是无关紧要的最后一个理由是，如果噪声确实有影响，我们可以做一个较好的同步器来解决，但谁也没有这样作。

当噪声是中等幅度时，由于会使系统进入 τ_{switch} 并不那么短的区间，从而确实破坏了同步器的时序。在噪声的幅度足够大时，就得到 Glasser 的理想判断器——这种判断器是完全随机的。在判断器从不观察输入数据的条件下，它就不需要花费无限长的时间去作判断，在这种情况下，就可以在很短时间内作出判断了。

6.6 全局时钟定时问题

解决全局时钟定时问题的正确方案与整个系统的设计有关，我们讨论几种常见情况。当系统中的某一块与其余部分有主从关系时最为简单，用片内时钟发生器产生一个“调谐”信号，系统的其余部分必然会随之响应。另一个相当简单的例子是片外产生一个良好的全摆幅时钟，这个系统时钟必须与系统中所有 VLSI 芯片用的时钟是同一类型的（例如，两相非重叠时钟）。由于不需要片内时钟缓冲器，芯片之间的时钟歪斜就减至最小，但是，对一个高性能系统来说，各相时钟之间的歪斜积累起来可以成为时钟周期

中的一个大部分，因而仍然是个问题。一块电路板上时钟歪斜为 5ns 已是相当不错的了，一个有若干个机柜的大系统时钟歪斜达 50ns 是十分普遍的。需要注意的是问题不仅在于歪斜使时钟信号扩展，而且也同样使数据线上的信号扩展。总而言之，这是与数据信号有关的时钟时序问题，而这些数据信号确定了对数据线上电压的解释。另一个问题是压焊丝的电感¹⁾，当低电压（读“TTL”）时钟从外部引入时，问题最为严重。这时时钟必须缓冲，系统时序容限就降低了。常常用一个滞后自举缓冲器使时钟振幅尽可能早开始，图 6.56 就是这种滞后自举缓冲器。

通常指的芯片输出信号歪斜是相对于输入时钟而言的，在这种情况下把芯片划分为两种时序方式是有好处的，如图 6.57 所示。 ϕ_{external} 和 ϕ_{internal} 之间的歪斜一般是很大的，同时这种歪斜完全越出了芯片设计时速度的额定范围。另一方面，至少可以用一个相对于 ϕ_{external} 歪斜很小的信号来使输出信号大致上同步。时钟缓冲器必须驱动的负载不同也会使歪斜有差别。由于外部时钟必须驱动所有电容，必然带来大的缓冲器延时，使 ϕ_{internal} 与外部时钟之间的歪斜变大，这些缓冲器延时对芯片制造工艺是很敏感的。

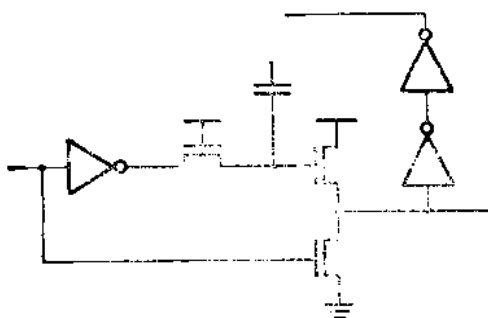


图 6.56 滞后自举缓冲器

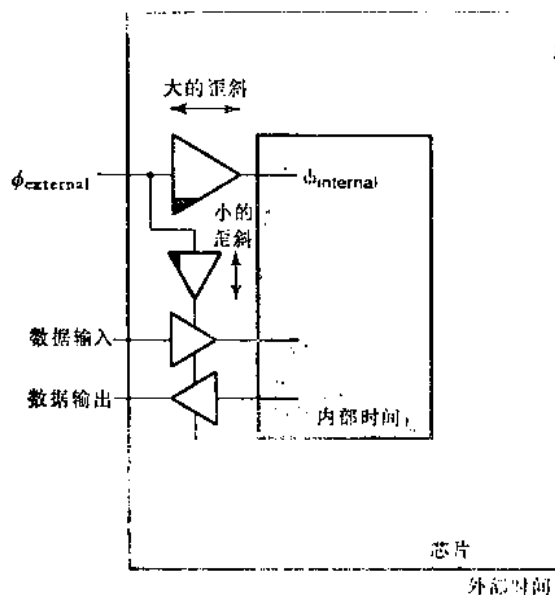


图 6.57 具有内部时间和外部时间时，使芯片时钟歪斜的接口电路

例 6.6 微处理器系统的定时。

图 6.58 是一个简单的微处理器系统方块图。微处理在 ϕ_1 时读寄存器，在 ϕ_2 时作 ALU 操作，并在 ϕ_3 写寄存器，I/O 操作可以选择在 ϕ_1 或在 ϕ_3 进行。如果我们选 ϕ_1 作 I/O 操作并又用了 ALU，则会浪费两个时钟节拍，另一方面，若选 ϕ_3 ，但使用寄存器而不用 ALU，则浪费掉一个时钟节拍。在本例中，我们已随意地选择了在 ϕ_3 进行 I/O 操作。

假定要产生一个与程序计数器（PC）具有某个偏离量的地址，其中 PC 和偏离量都存贮在寄存器中。为此需要作如下操作：

- 1) 电感问题也和产生片内时钟有关，这里指的是 V_{DD} 线的电感而不是时钟线的电感。地线上的 $L \frac{di}{dt}$ 经常是一个难题。

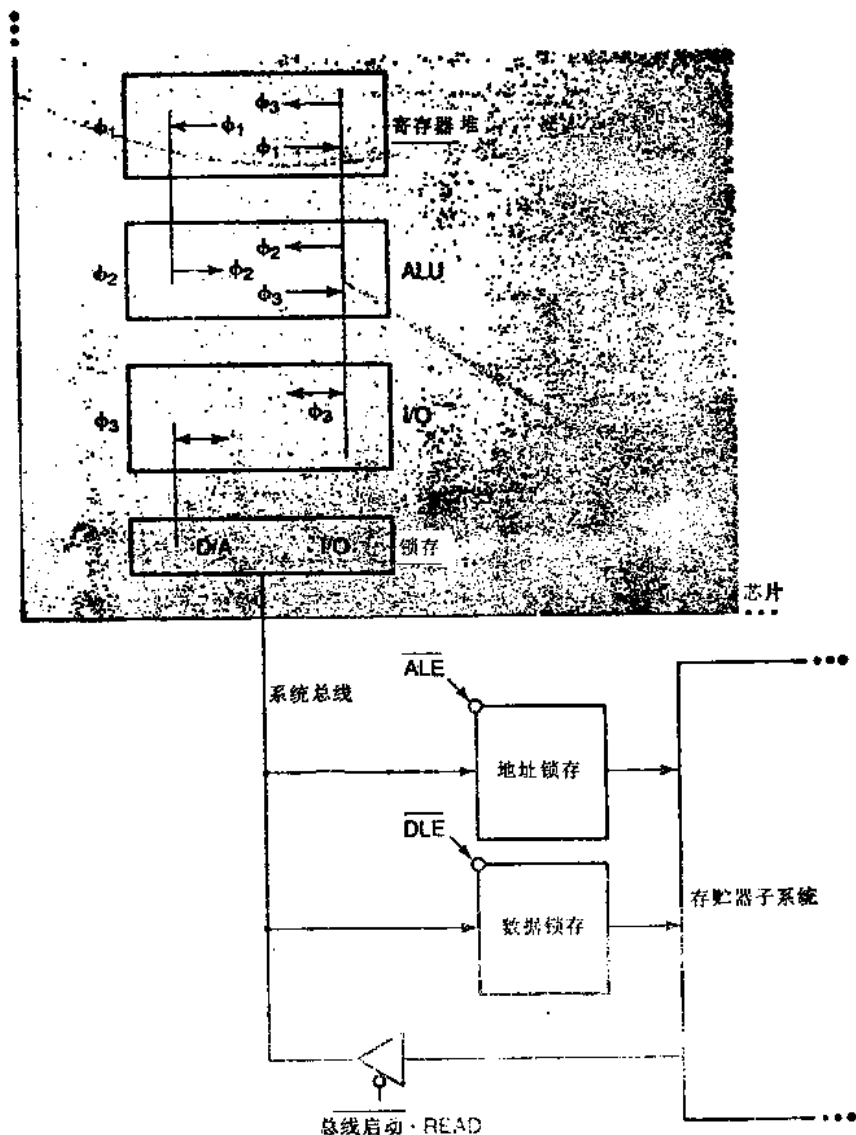


图6.58 由微处理器和外部存储器组成的系统的时序

- 同时从寄存器中读出 PC 及偏移量。
- 在 ALU 中相加以找出实际地址。
- 从 I/O 端口送出该地址。

以上操作分别在 ϕ_1 , ϕ_2 及 ϕ_3 完成。

假设系统有多路数据引出线和地址引出线，时序图示于图 6.59。与芯片外部的通讯过程如下：

- 主体（微处理器）在 ϕ_1 接近结束时使 $\overline{\text{BUS ENABLE}}$ 上升，即通知外部它需要占用母线，并给一个节拍时间使其它受控信号退出母线。
- 过一会，主体知道了要执行的任务（这时序不是十分苛刻的），就驱动 R/W 线对存储器进行读写。在所有受控信号退出母线的节拍期间，ALU 计算实际地址。

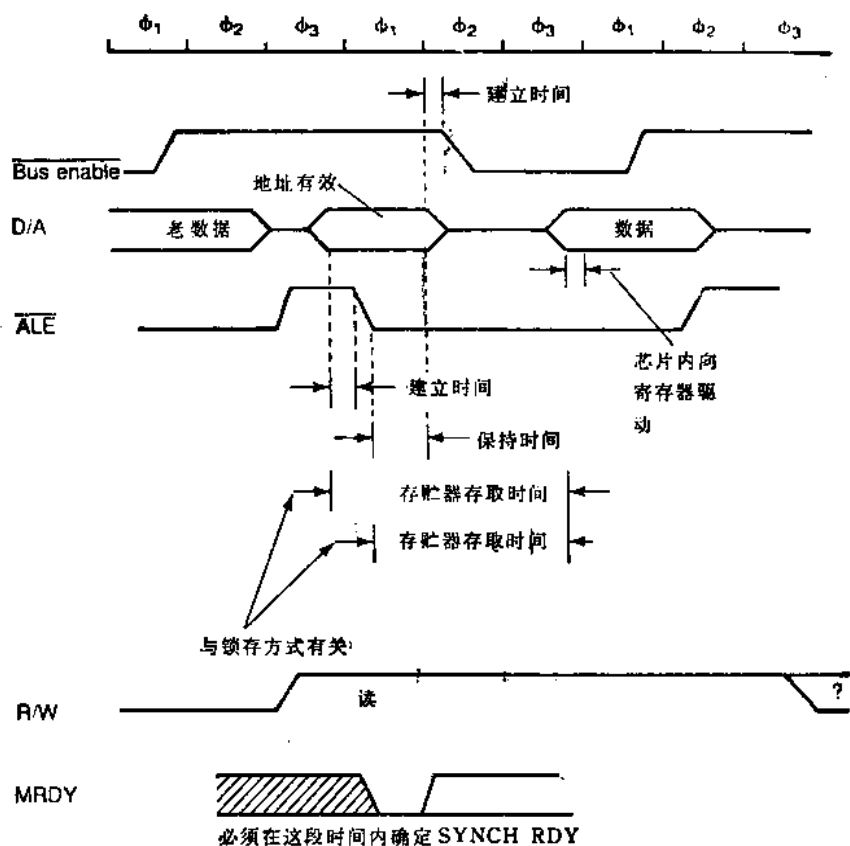


图6.59 微处理器及外部存储系统的时序

- 在 ϕ_1 期间地址被送入系统母线，如果外部地址锁存器是透明的，则当地址在系统母线上有效时，开始存取时间。
- 地址在系统母线上有效之后，主体等待一个建立时间，并且使地址锁存器启动信号 $\overline{\text{ALE}}$ 下降。建立时间是从地址有效到 $\overline{\text{ALE}}$ 离开有效“1”电平 (V_{OH}) 之间的时间。
- 在 ϕ_2 开始时，主体使系统母线驱动器变为三态，此时地址不再有效。从 $\overline{\text{ALE}}$ 降至有效低电平到地址第一次变为无效这段时间为保持时间。注意，在 $\overline{\text{ALE}}$ 正在下降的区间，既不是保持时间也不是建立时间，因而这段时间是浪费的。如果地址锁存器不是透明的，则存取时间从 $\overline{\text{ALE}}$ 达到有效“0”电平时开始。
- 母线驱动器变为高阻态之后， $\overline{\text{BUS ENABLE}}$ 变低以表示存储器现在已不能再与母线通讯了。由地址无效到 $\overline{\text{BUS ENABLE}}$ 离开高电平之间的建立时间一般为零。
- 在存储器进行工作和 ϕ_3 结束之前的某个时间，有效数据占用系统母线。 ϕ_3 的剩余时间留给微处理器把输入端的数据驱动到寄存器，该操作也许通过多路转换器进行（数据也可以保持在 I/O 锁存器，这意味着当下一次在 ALU 操作中使用这些数据时，应在 ϕ_1 时把它们驱动到 ALU，如同从寄存器驱动入 ALU 一样）。当数据在系统母线上有效时，存取时间就结束。

- 如果主体要进行另一个总线周期的工作, 则在下一次 ϕ_1 结束之前使 $\overline{\text{BUS ENABLE}}$ 和 ALE 上升。

若微处理器要对存储器进行写操作而不是读操作, 则另外还需要一个数据锁存启动信号 (DLE)。

在有异步信号条件下, 情况要更复杂一些。例如, 机器中可能有不同类型的地址——某些地址在高速存储器中, 另一些却在磁盘中。或者也可能有纠错码 (ECC), 于是没有错误的数据的存取很快, 而有错误的数据的存取则很慢。这样, 在正常情况下存取和已设计好的时序相一致, 但特殊情况下存取则需要一个随意的较长时间, 此时就需要增加一个 RDY 信号和一个同步器。

为了减少同步器失误的几率, 我们设计了一个存储器系统, 使得在存储器未准备好时保证 MRDY 线的信号为低电平, 但在任何时间它都能转为高电平。另外, 在正常存取的情况下, 要保证存储器能在 ϕ_2 之前确定它能符合正常的时序约束。例如, 存储器存取类型的确定是在类型变换 PLA 中进行的。如果地址在快速存储器中, 则在 ϕ_1 期间发出 MRDY 信号, 在 ϕ_1 结束时对 RDY 信号取样, 但在 ϕ_3 到来之前并不使用该信号, 这就使同步锁存器能有 1/3 或更多的微周期得以稳定。RDY 信号是在 ϕ_3 期间用来确定微周期是否应该继续下去或重复工作 (等待存储器作好准备)。这种处理方式示于图 6.60。

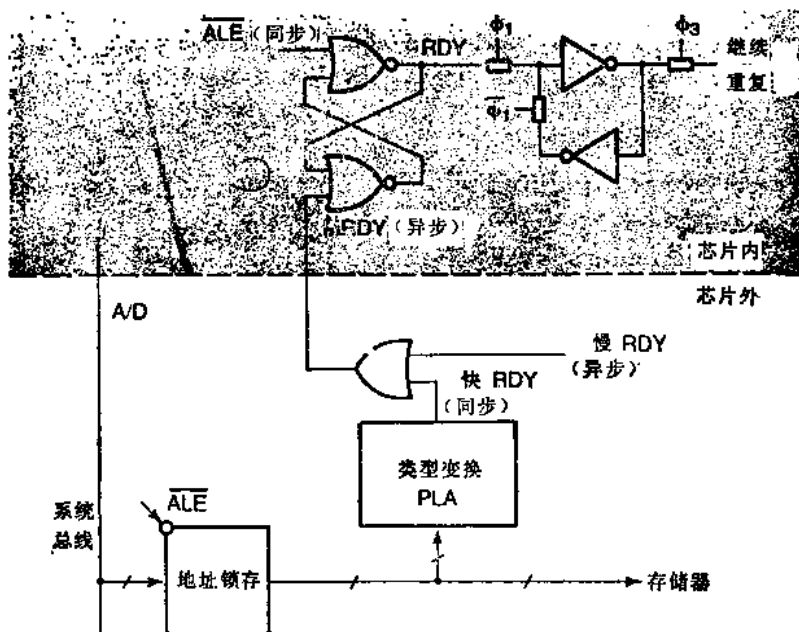
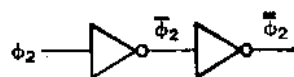
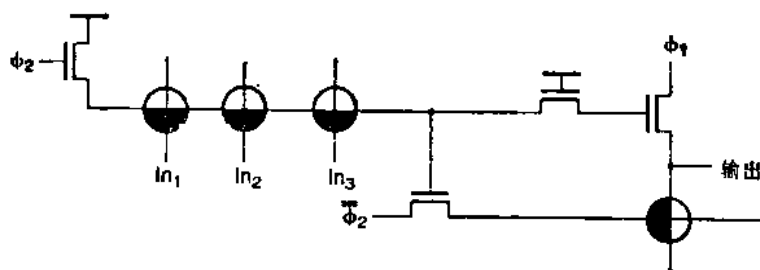


图 6.60 异步事件处理方式。图中大部分基准是同步的

6.7 小 结

信号的解释或意义与其在时间和空间中的相对位置有关, 本章中我们的注意力集中在时间轴上。一个数字信号通过一系列交替的时间作循环变化, 一会正确, 一会不正确,



6.63

图 6.64

.54 证明图6.65的锁存电路比图6.66的优越。

参 考 文 献

- [Anceau 82] F. Anceau, "A Synchronous Approach for Clocking VLSI Systems," *IEEE J. Solid-State Circuits* SC-17, 51-56, 1982. (On-chip phase-lock loop to eliminate skew due to clock buffers)
- [Bazes 83] M. Bazes, J. Nadir, D. Perlmutter, B. Mantel, and O. Zak, "A Programmable NMOS DRAM Controller for Microcomputer Systems with Dual-Port Memory and Error Checking and Correction," *IEEE J. Solid-State Circuits* SC-18, 164-172, 1983. (Single-phase clocking and distributed TTL input buffers)
- [Boysel 70] L.L. Boyse and J.P. Murphy, "Four phase LSI logic offers new approach to computer design," *Computer Design*, 141-146, April, 1970.
- [Couranz 75] G.R. Couranz and D.F. Wann, "Theoretical and Experimental Behavior of Synchronizers in the Metastable Region," *IEEE Trans. Computers* C-24, 604-616, 1975.
- [Ellul 75] J.P. Ellul, M.A. Copeland, and C.H. Chan, "MOS Capacitor Pull-up Circuits for High-Speed Dynamic Logic," *IEEE J. Solid-State Circuits* SC-10, 298-307, 1975.
- [Gardner 79] F.M. Gardner, *Phaselock Techniques*, Wiley, New York, 1979.
- [Goncalves 83] N.F. Goncalves and H.J. De Man, "NORA, A Racefree Dynamic CMOS Technique for Pipelined Logic Structures," *IEEE J. Solid-State Circuits* SC-18, 261-266, 1983.
- [Heller 84] L.G. Heller and W.R. Griffin, "Cascade Voltage Switch Logic, A Differential CMOS Logic Family," *IEEE International Solid-State Circuits Conf.*, 16-17, San Francisco, Calif., 1984.
- [Kinniment 76] D.J. Kinniment and J.V. Woods, "Synchronization and Arbitration Circuit in Digital Systems," *Proc. IEEE* 123, 961-966, 1976.
- [Krambeck 82] R.H. Krambeck, C.M. Lee, and H.-F.S. Law, "High-Speed Compact Circuits with CMOS," *IEEE J. Solid-State Circuits* SC-17, 614-619, 1982. (Domino CMOS)
- [Kung 82] R.I. Kung, S.T. Flannagan, and J.N. Spitz, "An 8K $\times$ 8 Dynamic RAM with Self-Refresh," *IEEE J. Solid-State Circuits* SC-17, 863-871, 1982. (Uses an arbiter)
- [Man 78] H.J. De Man, C.J. Vanderbulcke, and M.M. Van Cappellen, "High-Speed NNOS Circuits for ROM-Accumulator and Multiplier Type Digital Filters," *IEEE J. Solid-State Circuits* SC-13, 565-572, 1978.
- [Mead 80] C.A. Mead and L. Conway, *Introduction to VLSI Systems*, Addison-Wesley, Reading, Mass., 1980.
- [Suzuji 73] Y. Suzuji, K. Odagawa, and T. Abe, "Clocked CMOS Calculator Circuitry," *IEEE J. Solid-State Circuits* SC-8, 462-469, 1973.
- [Veendrick 80] H.J. M. Veendrick, "The Behavior of Flip-Flops Used as Synchronizers and Prediction of Their Failure Rate," *IEEE J. Solid-State Circuits* SC-15, 169-176, 1980.
- [Ward 83] S. Ward, private communication, 1983.
- [Watkins 67] B.G. Watkins, "A Low-Power Multiphase Circuit Technique," *IEEE J. Solid-State Circuits* SC-2, 213-220, 1967.

第七章 阵列结构的电路技术

我们在第五章中研究了如驱动器，读出放大器等专业小规模集成电路的设计问题。在第六章中，我们又讨论了时钟技术。本章，我们把这些集中起来并应用于设计大型的规则结构。

7.1 译 码 器

在规则的组合逻辑结构中，最简单的要算译码器了，因为它只有一级逻辑。我们在第一章中介绍过，译码器可以用几种不同的方法实现。以简单的静态方式而言， n MOS工艺的“或非”形式性能优于“与非”形式，这是因为下拉管是并联的而不是串联的。对于图7.1的有比CMOS电路，栅接地的 p 沟器件是一种耗尽型负载，用这种负载作为上拉器件，上述讨论照样适用。对于经典的CMOS电路，“与非”形式更为可取，这是因为“与非”形式允许晶体管串接，从而可以使用性能更好的 n 沟器件，这就是说，从电路性能受限制的角度考虑，人们宁愿使用五个串联的 n 沟器件而不想用五个串联的 p 沟器件。应当指出，一般总可以用把功能分解为若干个低扇入门级联的方法使逻辑门的扇入数减少到2，即使扇入数为5，这样处理也许仍然比只采用一个逻辑门多少要好一些。

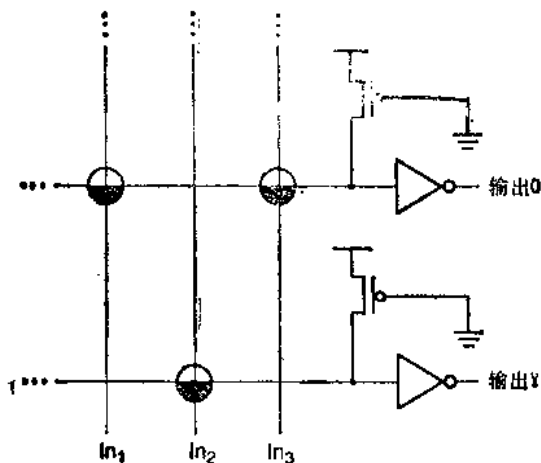


图7.1 有比CMOS“或非”型译码器

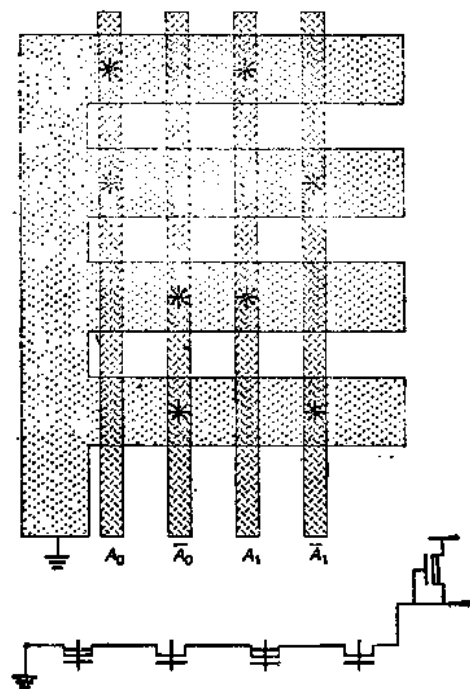
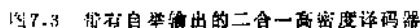


图7.2 离子注入编程“与非”型译码器

在一切阵列结构中，由于许多连线上的噪声同时在同一方向上变化形成的相干效应，使耦合噪声成为主要问题。于是，许多小噪声源在随机逻辑中或许不会造成困难，



但在高度密集的阵列中却是需要认真对待的。另一方面，因为规则阵列中的耦合作用是可以预测的，这就可以对噪声加以分析，用电路技术控制其幅度（例如对电压摆幅和预充电流加以限制，或产生差分信号），或者最起码预测所要求的噪声容限。既然能在噪声容限和速度之间进行折衷，能够估计噪声容限，就朝着提高速度前进了一大步。相邻连线间的耦合，线的交叉群间的耦合，以及多晶硅和扩散线的电阻都是至关重要的，预充造成的电流尖峰也是很重要的。

图7.3的电路显示了许多噪声问题，大多数问题集中在自举驱动级的耐噪声强度不足这一点上。由于预充线（ C_1 ， C_2 ，等等）跟随着耐噪声强度差的电路，从噪声观点来看此种电路是欠佳的。对 C_1 的预充受 M_6 控制，为此，我们首先研究 M_6 应当自举时的情况。 M_6 要自举， C_1 上的电压值必须高于 V_{DD} 与 M_7 的阈电压之差。然而，要注意如果 M_6 和 M_7 的大小一样，它们的阈电压值就是相同的，这样就更糟了，因为增强型n沟器件的上拉能力是如此之差，以致于 M_6 不大可能把 C_1 的电平一直上拉到 M_7 关闭。

存在着若干种解决这一困难的可能方案。一种方案是把 M_7 设计得比 M_6 长一些，从而使 M_7 的阈值高于 M_6 的阈值，甚至索性使 M_6 为零阈值器件。还有一种方案是恰好在 ϕ_2 变高之前使 V_{DD} 突然下降。应当记住噪声源之一是 V_{DD} 的跳动或“摆动”，假如刚好在 ϕ_2 变高之前 V_{DD} 反跳为高，则因 M_7 的导通会使自举作用丧失。还应当注意，甚至于预充线 ϕ_1 的下降也是有损于而不是有助于自举作用的。然而，更严重的问题是在 ϕ_1 下降之后大量地址线的变动状态，地址线下降时将会把 C_1 和 C_2 拉低，这又是另一种破坏自举作用的原因。这样，在 ϕ_1 降低之前使地址线稳定是很重要的（在本电路中，每条地址线与其补码配成对，如果版图上电容设计成对称的，则这种效应要比其它条件下弱，当然这种对称性在即使对中有偏差的条件下也必须成立）。

到此为止，全是在讨论在应当有自举作用时确保它能起这种作用的问题。第二种主要的情况是为了使输出保持低电平，在不应当有自举作用时就应该避免产生这种作用。地址译码器（如图7.3中的译码器）常常驱动预充电路，所以输出没有反跳是很重要的。这就要求 M_7 足够宽，使 M_6 的栅/漏覆盖电容不会造成 M_6 的栅电位上升到使其导通。这一点对该种译码器特别重要，因为它没有使 M_6 的源保持下降的晶体管，因此，如果源电压一直上升，就将不可恢复。

对CMOS译码器，可把反相器用作缓冲器，也可采用图7.4的预充电路（注意逻辑变换）。在这一电路中， $NR_{IT} = |V_{TP}|$ ，所以在图7.3的电路中所有的噪声问题在本电路

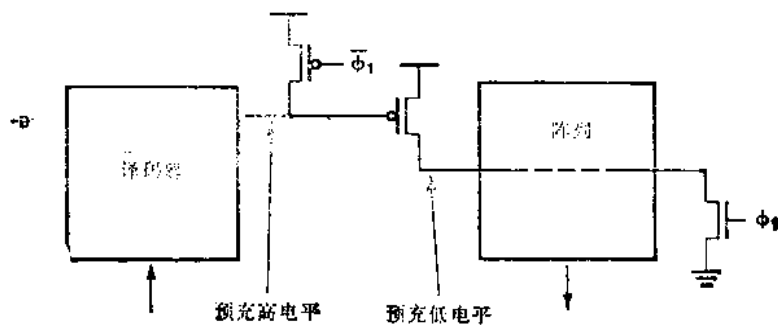


图7.4 预充CMOS译码器

中也必须加以考虑。图7.5为另一种译码器，其设计方式与图7.3中的相同，增加的译码功能在输出缓冲器中完成。

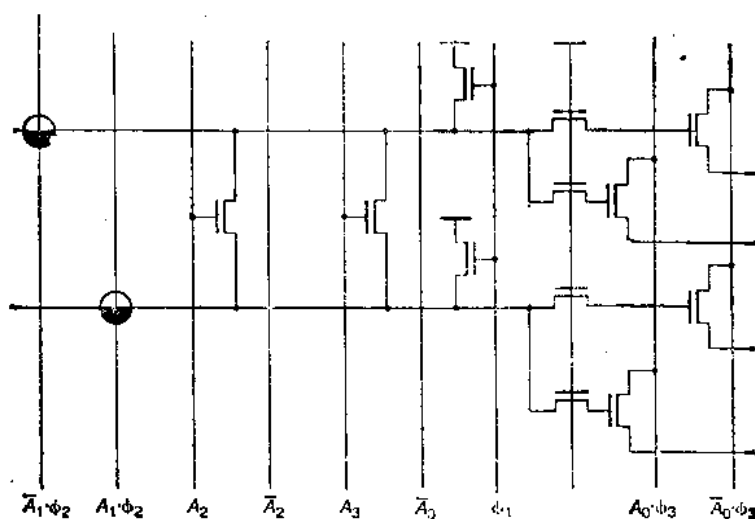


图7.5 四合一高密度译码器

7.2 ROM和PLA

ROM 和 PLA是两种最常见的用于组合逻辑的阵列结构。在研究这些阵列时我们将会看到许多不同抽象领域的相互作用。

ROM的设计问题

在第二章的最后部分，我们对 HROM 作了详细的分析和模拟，在那里我们面临的电路问题是多晶硅字线传播时间和地线的噪声。在第四章，我们又回到了这个难题，并且搞清了位线之间的耦合是主要的问题。接着，我们在第五章分析了读出放大器适合于放在 ROM 的输出端，但我们并没有结束这组命题。在本节，我们再研究三个问题：虚接地的利用，XROM 版图设计技术以及逻辑优化。

我们在译码器设计时用过的虚接地技术可以用于 ROM 位线，如图7.6所示，利用使一条字线上升和用 Y_i 线来选择正确的位来实现对 ROM 中晶体管的访问。例如，用 Y_i 上升访问 B_i 列的晶体管，其输出被送到 z_i ，同时，由选择器对 B_i 列编程，并送到 z_0 。所有列在 ϕ 期间预充到高电平。在虚接地 ROM 电路中，每一位平均需要一条水平引线和一条垂直引线，而对非虚接地电路，由于一定是共享地线，每一位要求有 $1 - \frac{1}{2} \times 1$ 条线。

虚接地 ROM 的主要问题是潜通路。例如， Y_i 和 W_i 上升会使这一组中的所有位线放电，这就是造成输出选择电路复杂化的原因。 B_0 和 B_i 之间留的空间就是为了阻挡潜通路，如果没有这段空间，某一组中一条位线下降会由于潜通路而使邻近组的输出下降，但是，为了做到这一点，可能要用许多串联晶体管把相邻的输出下拉，所以，与其在阵

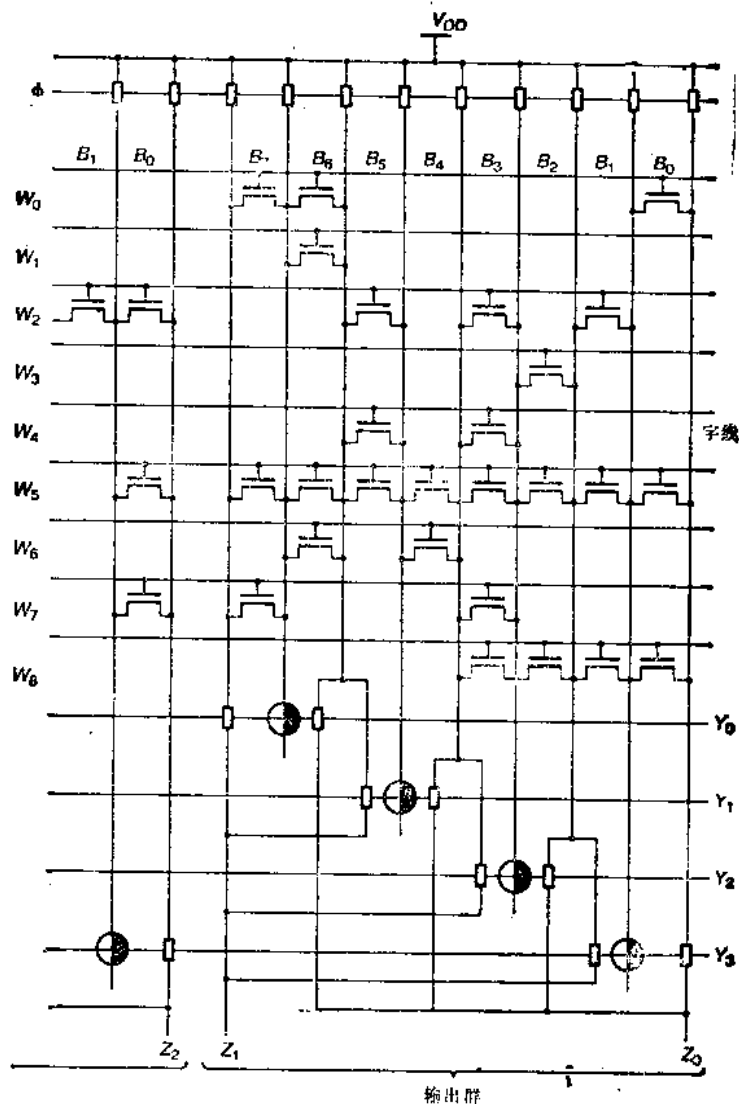


图7.6 虚接地ROM的一部分

列中插入空间以制止这种效应，不如用与预充器件并联的高阻上拉管，或者用一种在某个时间之后能暂时停止输出的电路。由于潜通路的下拉作用比正常通道慢得多，只有较快的跳变到来时才能选中所要的通道。

增强型器件当然可以用于预充，但用耗尽型器件使连到读出放大器上的线上拉也是合理的。这些耗尽型器件一直把电平上拉到 V_{DD} ，而唯一的缺点是在位线下降时，耗尽器件反过来却导通了。不过，对于大多数合理设计的电路，在上拉管转导通以前位线就已经被读出了。

另一种可能的方案是预充低电平而不是预充高电平。可以预期，一条高电位的字线会使一个晶体管接通，而该晶体管的作用则是使位线拉高。我们知道上拉后期电平变化比电路设计成下拉的情况慢得多，但输出也许在开始的半伏电压内就会被读出，同时起始摆率也是差不多的。

当地线与位线平行或采用虚接地技术时,有一种版图设计窍门可使 ROM 的单元面积大大减小,这种技术示于图7.7,得到的有源区图形呈“X”形而不是“H”形,因此自然而然地称之为 XROM。由于版图的高度紧凑和晶体管的奇特形状,使这种 XROM 的分析相当复杂。一般说, XROM 单元中的晶体管很窄,其阈值比电路中采用的大部分其它晶体管的阈值高得多,这主要是由场注入杂质在沟道下的侧向扩散造成的,因此,这些晶体管还有另外的特征。XROM 常常采用虚接地技术设计,电路速度一般约为 HROM 的一半。

用离子注入编码方法(如图7.2的译码器),或采用每个交点多于一位的编码设计,可进一步取得更高的版图密度,这是借助于对沟道注入和不同晶体管尺寸得到的附加信息编码来实现的。这种 ROM 的读放电路对每个节点进行多位编码,必定是相当复杂的。

研究不同的版图尺寸可以获得优化设计。ROM 的优点之一是可以行数和列数之间进行折衷,即长宽比是可调的。如果把行数缩小到16或32,就会发现大部分列不是全部被用完了就是完全不必编程了,这两者中的任一种,特别对第二种情况,就可直接取消该列,从而缩小 ROM 的面积[Gutttag 82]。

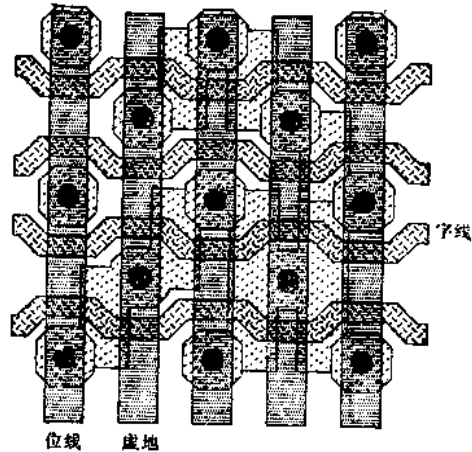


图7.7 XROM版图

PLA的设计问题

可编程逻辑阵列是和 ROM 紧密相关的,并有许多种合并形成的或派生的结构。虚接地技术在纯粹的 PLA 中不能工作,读出放大器设计和噪声计算也由于若干行可以同时被选中而更加困难。另一方面,也有一些 ROM 中不存在的自由度,其中大部分表现在逻辑和拓扑设计上。

多数大的阵列是要预充的,而 PLA 也无例外。图7.8和图7.9为两种预充技术,这两种 PLA 都采用两相非重叠时钟。一种单相 PLA 示于图1.70中。图7.8的电路采用了虚接地技术,使输入稳定之前阵列不进行取值。在 ϕ_1 期间把“与”阵列预充成高电平,尽管这时输入端电平可能会变化,但“与”阵列输出端不存在对地的通路,从而不会放电。这“与”阵列在 ϕ_2 开始上升之前必需稳定,当 ϕ_2 一上升,所有“与”阵列中的虚接地线就要下降,这就会在阵列的内外引起大量噪声。外部噪声由地线上电流波动造成,而内部噪声则是由虚接地线的下降引起;虚接地线通过电容把输入端下拉(可能电路变慢),同时也使预充的“与”阵列输出端下拉。因此,把图7.3那种自举缓冲器放在“与”阵列和“或”阵列之间是十分危险的。在 ϕ_2 期间“与”阵列取值以及预充“或”阵列,而在 ϕ_1 期间在对“与”阵列进行预充的同时“或”阵列取值。此时,可以把“或”阵列的输出引入“与”阵列而形成有限状态机。

图7.9是第二种两相 PLA 的略图。“与”阵列在 ϕ_2 时预充,同时“或”阵列取值; ϕ_1 时则“或”阵列预充而“与”阵列取值。这种 PLA 中不用虚接地技术,在预充阵列的同时与

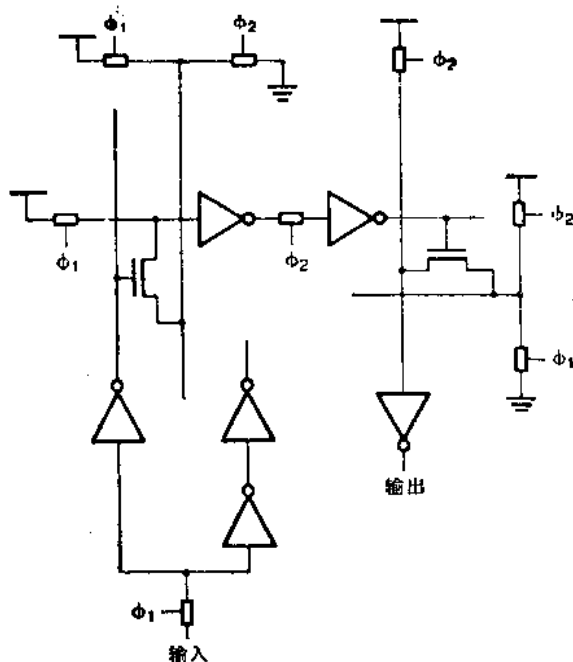


图7.8 虚接地PLA

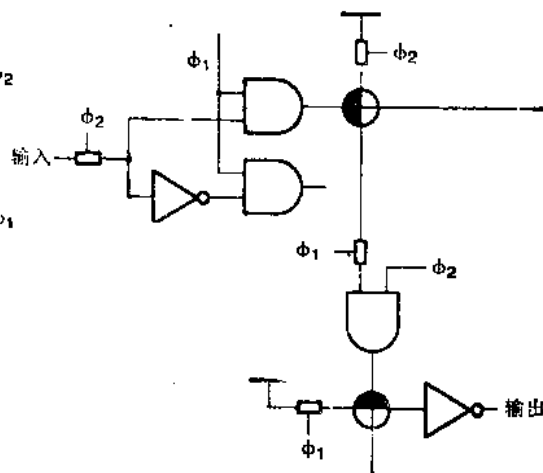


图7.9 预充两相PLA

下拉晶体管相连的输入端保持低电平。

在 nMOS 电路中，可能会用时钟控制的自举缓冲器来实现“与门”电路。这些缓冲器设计得最紧凑时[例如 图 5.12(e)]，可使输出预充到低电平，但并不予以保持，因此，检查和该输出线的耦合是十分重要的。存在一种可能性，即一条字线的两端都有一个向高电平上升的“或”阵列，而该字线应当处于低电平，这样一来就会把低电平的字线向高拉。另一方面，“或”阵列（应该认为有许多这种“或”阵列）的位线下降将会把字线电平往下推而有助于它们保持低电平。一般情况下，这第二种影响起主要作用。在“与”阵列中这第二种影响肯定起主要作用，因为即使对于任何随随便便的设计，除了极少数乘积项外都将是低电平。值得注意的是时钟线上的电容负载是和 PLA 编程有关的。

为了避免需要用一整个节拍时间来进行预充，常采用闪烁预充时钟。这种时钟往往在时钟脉冲起始附近出现，并只持续整个节拍时间的一小部分，所以它被用于预充在该节拍后取值的那些元件。

PLA 可以从很多方面进行优化，下面我们将分析几个例子。最明显的方法是直接的逻辑最小化。即使对于中等规模的输入和输出矢量，从计算角度看乘积项(行)数量的最小化¹⁾也是难以对付的。但是，开发了许多直接推断的技术来使乘积项的数目有所减少(10到30%)[Hong 74]。

例如，下面两个乘积之和的表示式：

$$X = \overline{A}\overline{B}\overline{C} + \overline{A}B\overline{C} + \overline{A}BC + ABC \quad (7.1)$$

以及

$$Y = \overline{A}B\overline{C} + \overline{A}\overline{B}C + A\overline{B}C \quad (7.2)$$

1) 这里使用的是Quine-McCluskey方法。

代表了八种可能的乘积项中的六种，可以把它们简化为

$$X = \overline{A}\overline{B}\overline{C} + \overline{A}B + BC \quad (7.3)$$

以及

$$Y = \overline{A}\overline{B}\overline{C} + \overline{B}C \quad (7.4)$$

因此只用了四种乘积项就实现了方程 (7.3) 和 (7.4)。把 X 改写为：

$$X = \overline{A}\overline{B}\overline{C} + \overline{A}B + ABC \quad (7.5)$$

就需要多用一个晶体管，但因乘积项数目没有变化而保持阵列的大小不变。

用一般方法构成“与”阵列，晶体管的利用率是不高的。以两输入串 A 和 B 为例，若这些输入串互相交替，则对 PLA “与”阵列的输入为

$$(a_0, \overline{a}_0, b_0, \overline{b}_0, a_1, \overline{a}_1, b_1, \overline{b}_1, a_2, \dots) \quad (7.6)$$

所以每位有四个输入端，并且表示为

$$(a_i, \overline{a}_i, b_i, \overline{b}_i) \quad (7.7)$$

对于每两个变量 (a_i 和 b_i)，在“与”阵列中每个乘积项有四个放晶体管的位置，这就意味着存在16种可能的编程组合，也说明了每两个变量有16种可能的逻辑功能。这是一种惊人的巧合，但不幸的是当以方程 (7.7) 的形式把输入放到阵列中时，它的好处并不能得到，这是因为有许多种编程组合是等效的，例如， $a_i + \overline{a}_i$ 的作用是和 $b_i + \overline{b}_i$ 一样的(扫兴!)，而我们本来想要在阵列中引入 a_i 和 b_i 四个函数以便在每一个乘积项实现16种功能的编程。引入阵列的正确的函数是 Karnaugh 图的元素，即为

$$(a_i b_i, a_i \overline{b}_i, \overline{a}_i b_i, \overline{a}_i \overline{b}_i) \quad (7.8)$$

要做到这一点就需要在阵列的输入端放置逻辑门而不是简单的反相器。所有以前用单个乘积项实现的功能现在仍然可以用单个乘积项得到，此外，还有整整一类有用的函数过去是很难实现的，但现在都可以用单个乘积项来做了。例如，PLA 不善于检查两个输入串 A 和 B 是否相等是众所周知的，然而用编程输入，因为实现像 $a_i \overline{b}_i + \overline{a}_i b_i$ 这种函数容易得多，这种检验就可用单个乘积项来完成。但是要注意， a_i 是作为 $a_i b_i + a_i \overline{b}_i$ 以及 $a_i + \overline{b}_i$ 是作为 $a_i b_i + a_i \overline{b}_i + \overline{a}_i \overline{b}_i$ 执行的。

优化的另一个方面是 PLA 的分解和折叠。我们取一种如图7.10所示的标准的 PLA 来讨论折叠技术，首先去掉无用部分以减少其尺寸，然后用不同生成方法对阵列进行折

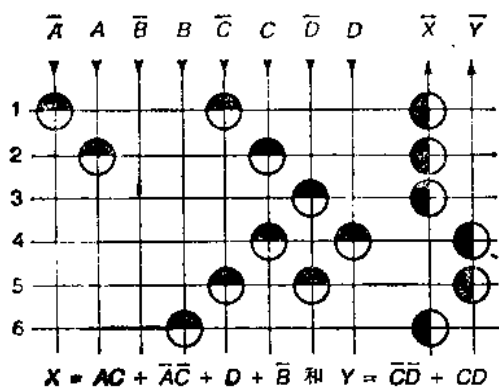


图7.10 执行 $X = AC + \overline{A}\overline{C} + \overline{B} + D$ 和 $Y = \overline{C}\overline{D} + CD$ 功能的PLA[Hachtel 82]

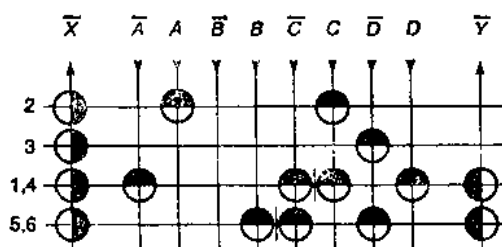


图7.11 用行折叠改进图7.10中的PLA[Hachtel 82]

叠来填满空间。两种折叠阵列示于图7.11和图7.12[Hachtel 82]，它们的功能和图7.10

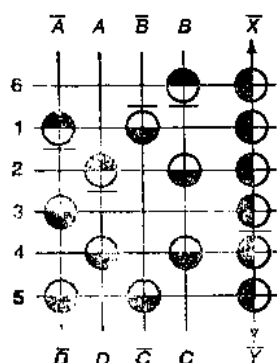


图7.12 用列折叠改进图7.10中的PLA[Hachtel 82]

中的 PLA 是完全一样的。这种技术的主要缺点是输入和输出的位置分散了，也许不利于布局，也很难实现。注意，折叠可以和输入编码结合起来进行。

存储逻辑阵列 (SLA) 设计对布局十分注意 [Patil 79]。PLA 之所以能折叠是因为一般情况下它们的布局是不紧凑的，特别当 PLA 是以算法从较高级的表象用计算机程序产生时更是如此。在大部分巨系统中，通讯是局部的，用这种系统得到的设计，只有极少数连线是全局的，所以大多数连线只连接到少数几个地方（如果每一个门都和其它所有门一一相连，则具有 100000 门的芯片看起来将会是怎样的呢？）。这就是 PLA 布局不紧凑的部分原因。

所以，人们希望像我们在折叠过程中所做的那样，通过重新安排 PLA 的各部分，得到一种没有长连线的规则结构，特别是“或”阵列的输出可以和“与”阵列的输入交替。这就是 PLA 的基本原理。一般情况下，为了提供更大的设计自由度，寄存器和逻辑阵列中分裂出来的一些部分交替的。

PLA 的稀疏特性从电路的角度看是可以利用的。对 PLA(或 ROM) 速度限制的因素之一是放置于各交叉点的晶体管的尺寸。如果 PLA 的布局是松散的，并且又有一种巧妙的重新安排方法，则只要某个阵列位置邻近的交叉点没有编程，该位置上放置的晶体管就可以比一个阵列点所占空间大。在少数重新安排不成功的情况下，就只好扩大阵列了。

我们已经研究了几种 PLA 的优化方法，毫无疑问，在逻辑阵列方面所做的优秀工作还多得多[例如，Weinberger 67, Lopez 80]，我们的主要目的在于向读者提供已经证明是有益的思考方法。应当指出，版图，拓扑，电路，逻辑等等方面都是有作用的，而且并非互不相干的，某些方面进行了优化会影响到另一些方面优化的潜力。正如本文所述，这诸方面如何相互作用是一个尚未完全解决的问题。已经有了一些适用于个别领域的优化程序，但是，覆盖多个方面的协调一致的优化技术还没有解决，从而给未来的研究人员和计算机提出了复杂的任务。

7.3 RAM

有多种随机存储器 (RAM) 可以用于 VLSI 芯片。一般说，存储单元中的晶体管越少，外围电路中的晶体管就越多，并且外围电路也越庞大。本节从外围电路最简单的存储单元开始讨论。

八晶体管RAM

最简单的 RAM 设计采用单母线和交叉耦合反相器，两种单元设计示于图7.13。图7.13(a) 为静态单元，并要求预充母线，写单元的时间包括了沿反相器链传播信号的时间，这个时间不会短，因为单元中使用的器件一般都很小，以便可以用位线来改写单

元,同时可以节省功耗和面积。图7.13(b)所示为半静态单元(只有在 ϕ_1 高时才能使时钟停止),这种单元比前一种少用一个晶体管,但需要增加一条控制线。

这些单元可以简单地用增加数据线和控制线的方法而成为多端口的,如图7.14所示。当然,在这种情况下单元必须驱动两根母线,因而使本来就驱动能力很弱的晶体管又增加了负担。

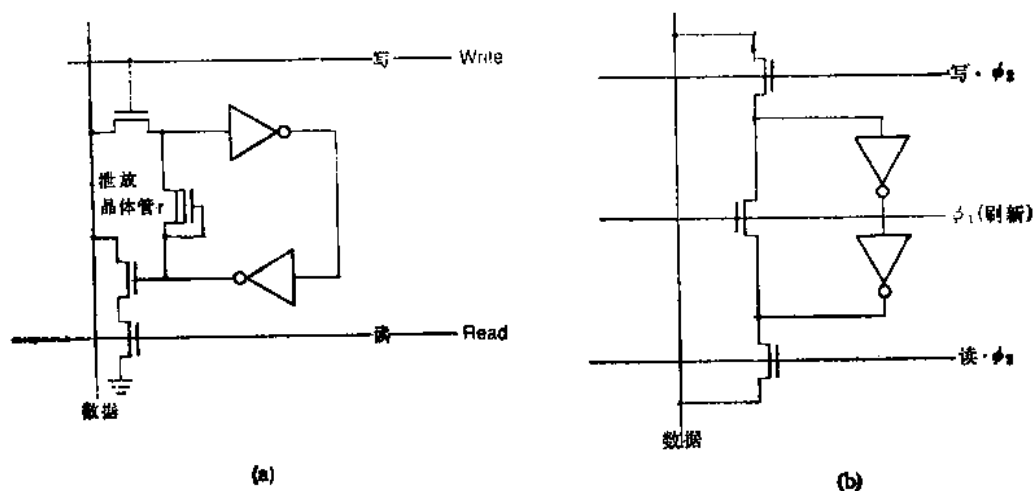


图7.13 八管RAM。(a)静态, b)半静态

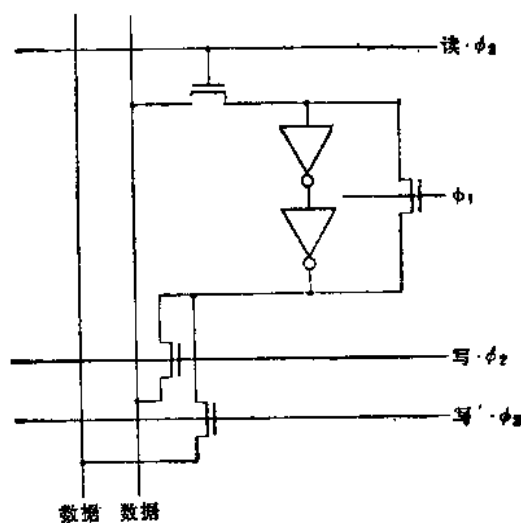


图7.14 双端口八管RAM

六晶体管RAM

图7.15为两种六管RAM单元,由于这些单元产生差分输出信号,它们在大阵列中使用也会相当的快。输出差分信号消除了各种共模噪声源,这就允许使用较小噪声容限的读出放大器,而较小的噪声容限又使更高的速度成为可能。另外,对于六管RAM来说,写单元的速度受写放大器中晶体管速度的限制,而不是像图7.13中的八管单元那样由单元中的晶体管速度决定。

写操作是用提升行选,以及降低一条位线同时提升另一条位线完成,这将迫使选中

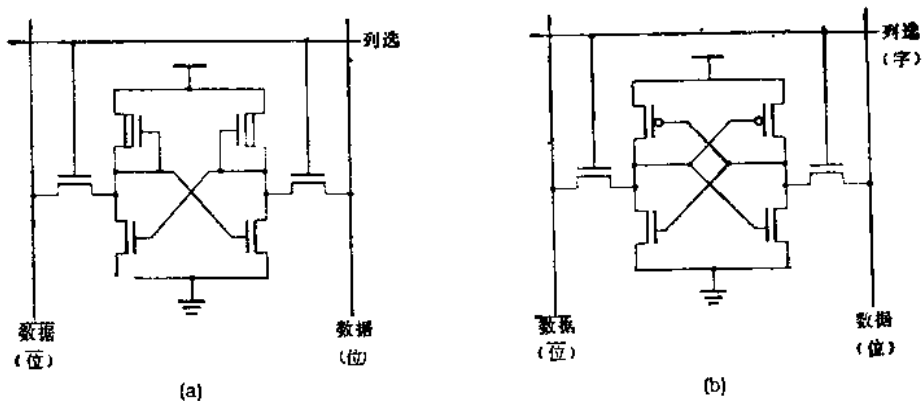


图7.15 六管RAM。(a) nMOS; (b) CMOS

单元的一边降低，达到存储数据的目的。读操作则是用同时提升两根位线为高并选择所要的字线，于是单元将把一条位线往低拉，当电平拉开到足够大时，差分读出放大器就工作了。典型的静态 RAM 的整体结构示于图7.16，当然还有许多种不同的形式。列选择器¹⁾相对于读出放大器的布局是一个重要的问题，另一个重要的问题是共享读出放大器的列数。在共享读出放大器时，差分输入信号的走线必须认真考虑，要做到对读出放大器来说噪声和负载电容都是一样的（实际上负载电容决不会完全一致，因为这些电容与阵列中的数据有点关系）。

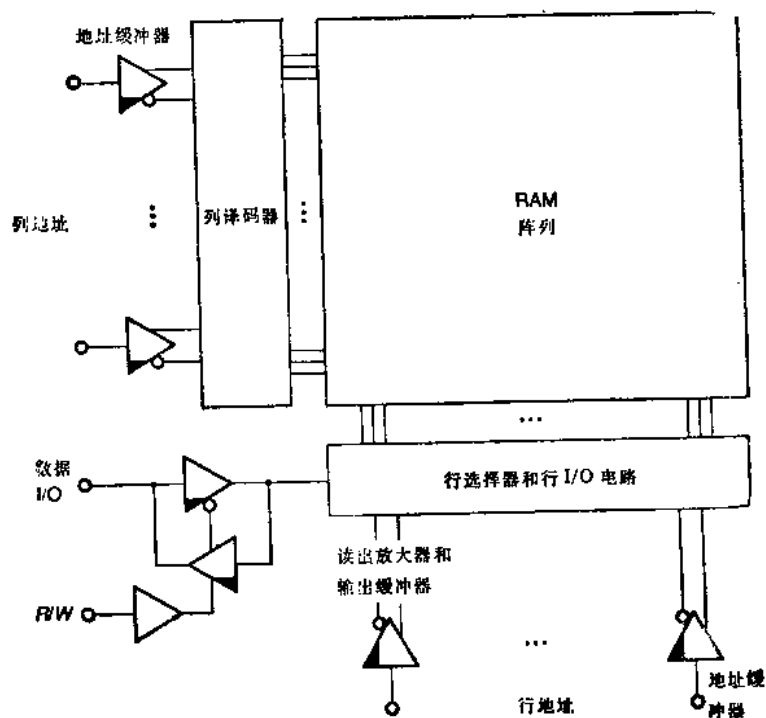


图7.16 RAM阵列的结构

1) 列选择器常常称为Y译码器，行选择器则叫做X译码器。

在 RAM 单元内部, 下拉晶体管的大小往往是传输晶体管的 2—3 倍, 对 nMOS 来说, 上拉晶体管存在功耗问题。上拉晶体管只需要补充因漏电及 α 粒子轰击引起的电流损失, 因此兆欧级比千欧级更好。实现这一点最廉价的方法是采用弱耗尽注入 (在我们的 $2\mu\text{m}$ 示范工艺中, 若在同一晶体管上用增强型注入和耗尽型注入就可以得到这种器件)。商品化的静态 RAM 芯片采用多晶硅电阻器, 其阻抗范围为 $\text{M}\Omega/\square$, 但这就要求用额外的工艺步骤。图 7.15(b) 中的 p 型上拉晶体管从功耗的角度看当然是很好的, 但器件的面积则成问题。我们在 5.4 节中讨论过, 为了防止单元中信息的丢失, 单元的环路增益不能低于 1。静态 RAM 在某一瞬间两条字线会同时打开, 从把位线拉低和环路增益减小来说, 这是一种最坏条件。如果上拉作用很弱, 当字线为高时, 单元 (nMOS 或 CMOS) 看起来就像交叉耦合的增强型负载 nMOS 反相器。

时钟的和非时钟的静态 RAM 设计的主要差别在于使用的读出放大器, 行驱动器, 以及位线上拉器件等类型的不同。在全静态 (非时钟) 条件下, 为了减少两个相继的读操作之间的恢复时间就要限制位线电压摆幅, 之所以采用增强型位线上拉器件, 已经证明, 时钟的静态 RAM 设计比非时钟的静态 RAM 设计性能更为优越。在位线上设置泄放下拉器件是很重要的, 这样就可以防止由次开启效应造成的, 在两次存取之间预充电电压随时间变化的不可控制。

由于采用差分方式来读单元, 我们最关心的是差分噪声, 邻近的位线是一个大的差分噪声源。读出放大器的速度受共模信号以及如差分负载电容这类把共模噪声转化为差分噪声作用的影响, 从这方面看, 共模噪声也是要重视的。其它差分噪声源有读 1 后跟着读 0 的图案效应, 以及读出放大器输入晶体管之间的失配等。

由于常常用扩散线做电源线, 在 RAM 阵列中 V_{DD} 和接地噪声是相当大的。对于某些 CMOS 设计, 实际上是多晶硅作 V_{DD} 线, 由读操作引起的电流尖峰表现为电压噪声。注意, 根据地线平行于字线或平行于位线的差别, 噪声的作用是很不一样的。在某种情况下噪声干扰了那些访问到的单元, 而在另一种情况下则是那些未访问到的单元受到干扰。正如我们在第二章中例举的 ROM 实例中看到的那样, 浪涌电流会造成地线电压大大增加, 这就要引起存取时间的增加。

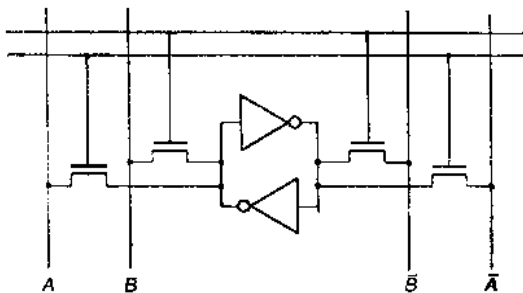


图 7.17 双端六管 RAM

双端六管 RAM 的设计如图 7.17 所

示。当单元中两对存贮晶体管都打开时, 双端操作使有效上拉电阻减半, 这就降低了交叉耦合反相器的环路增益并减少其抗干扰能力。

四晶体管和三晶体管 RAM

商品化的存贮器是用特殊工艺制造的。静态 RAM 工艺不同于动态 RAM 工艺, 而与逻辑电路的工艺十分类似, 然而, 静态 RAM 工艺常常有双层多晶硅和特殊的多晶硅电阻。把 RAM 设计为微处理器的一个部分时, 一般不选高阻多晶硅负载工艺, 但耗尽型或 p 型负载使用的面积, 或许还有消耗的功率都是惊人的, 另一种办法是干脆把上拉器

件放在电路外部，这样一来存储器就变成动态的了。Wooley 指出 [Wooley 80]，对于 ROM 设计来说，我们用动态技术来节省功耗，而在 RAM 设计中，采用动态技术不仅节省了芯片面积，而且也减小了功耗。

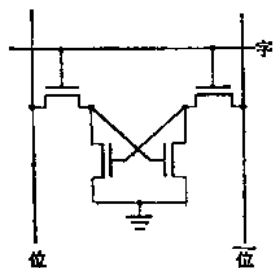


图 7.18 四管 RAM

图 7.18 为四管 RAM 单元，这种设计带来的新问题是需要对单元进行刷新（约每毫秒刷新一次），单元的噪声容限下降（高电平端只上拉到 $V_{DD} - V_T$ ），以及对单元回写时需要使位线恢复高电平。也就是说，必须以某种方式恢复单元的高电平，既然单元中没有上拉器件，那就一定要在外部对单元进行刷新。

例 7.1 用于移位寄存器的三管 RAM 单元。

Lyon 曾经用三管动态 RAM 单元制造了一个很长的移位寄存器并用于信号处理 [Lyon 80]。下面研究的系统由四个部分组成（不包括某些特殊的部分），其基本构造与 Lyon 的是一致的。

三管 RAM 单元是设计的核心，两种三管 RAM 单元示于图 7.19 和图 7.20。信息由控制 M_1 选通，动态地存储在源接地的晶体管的栅接点。位线一定要预充以执行一个读周期，进行读操作时打开 M_3 ，并且确定预充的位线是否已放电。这些单元由于寄生效应极为重要而相当难以设计的，主要问题是向 M_2 的栅存低电平，在读操作时与 M_2 耦合在一起的 C_{DC} 会促使该电平升高。当 M_2 低于阈值时，这个问题就特别突出，因为这时 M_2 的输入电容极小，不太大的耦合就会使该晶体管打开（应当注意，SPICE2G 5 对低于阈值的 C_{GS} 估计过大了，因而过高地估计了单元的鲁棒性）。

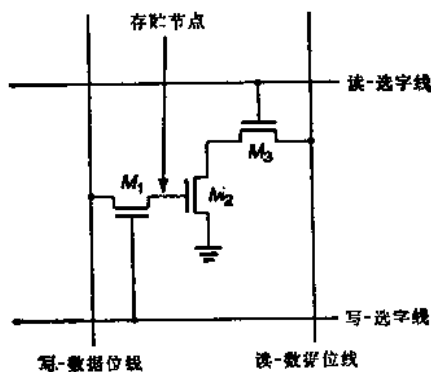


图 7.19 具有分开读/写位线的三管 RAM

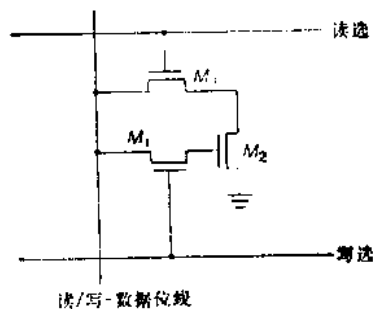


图 7.20 具有共用读/写位线的三管 RAM

为了构造一个长的移位寄存器，我们先用两个图 7.21 的 M 位串联移位寄存器作为输入和输出电路，在这两个串联移位寄存器之间的是一个 $M \times M$ 位三管动态 RAM (dRAM) 单元阵列。数据以每时钟周期一位的速率进入串联移位寄存器，在 dRAM 阵列中， M 位平行移位，其速率为位进入串联移位寄存器速率的 $1/M$ ，所以每个周期平行移位一列，这就保证了时序的正确。在 M 位进入阵列后他们就向 dRAM 阵列平行移位；列从右边开始向右移位，并且往左继续下去，直到全部列都移完；然后再在左边开始该过程。每一列在移位时进行刷新，系统的第四部分即为刷新电路。这种设计所用的面积和功耗比普通的 nMOS 移位寄存器设计小得多。

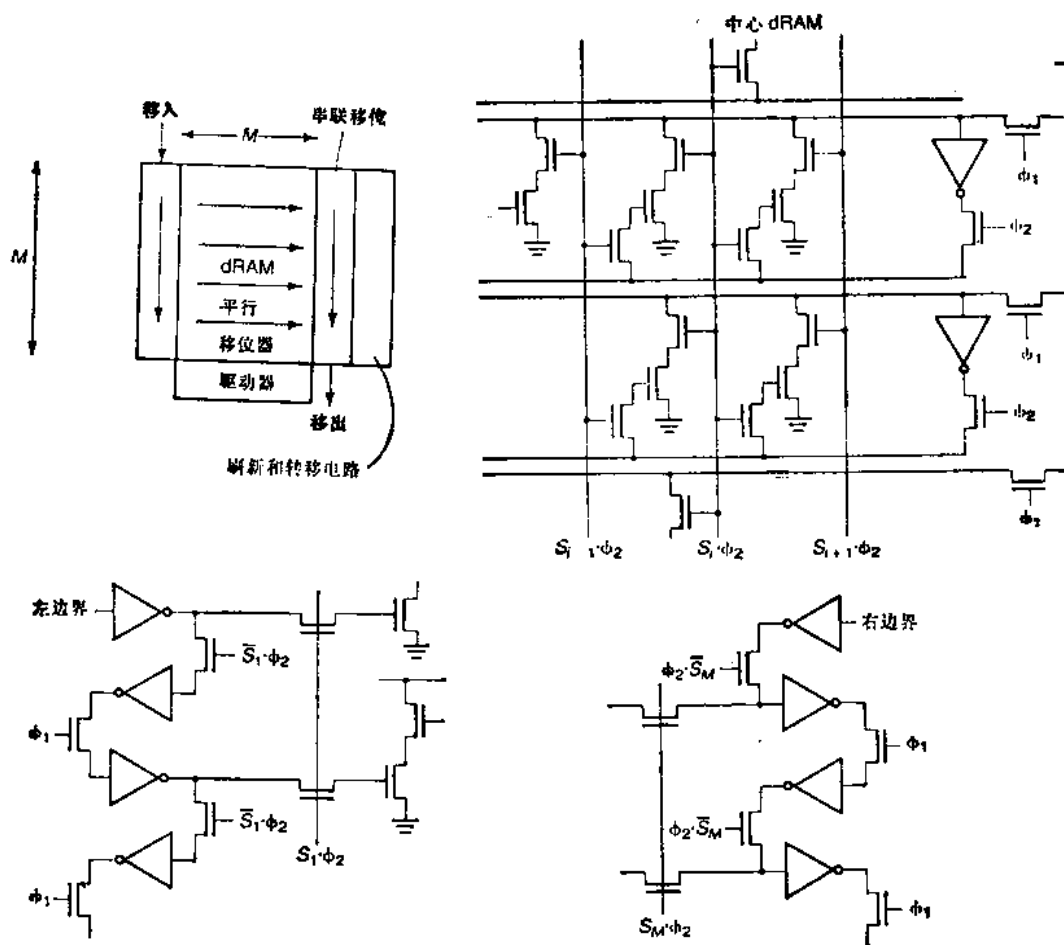


图7.21 基于三管RAM单元的移位寄存器的结构

单晶体管RAM

在存储器中，单管存储单元是最简单的同时又是最复杂的[Dennard 68]，它代表的也许是VLSI设计中最复杂的所在。虽然单管dRAM的细节问题并不是本书的范围，我们仍将向读者介绍有关的资料。

图7.22为单管dRAM的略图。商品化RAM的单元阵列所占面积通常小于50%，其余

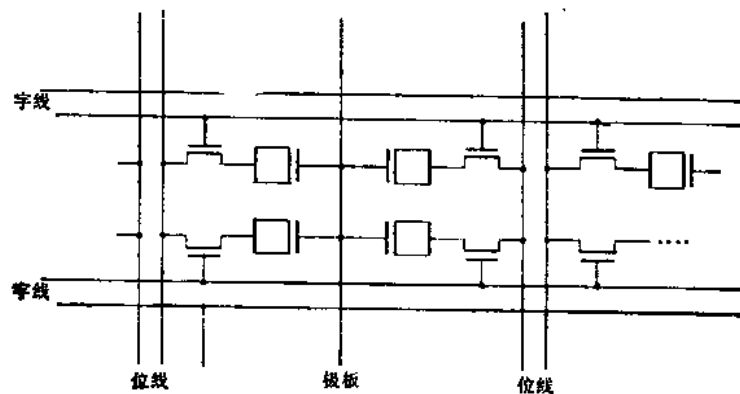


图7.22 单管RAM

的部分用于时钟、读出放大器、驱动器等等。单元版图、制造工艺、电路设计，以及阵列结构这四个方面的问题是紧密相关的。

单元本身由一个晶体管和一个电容器组成，以电荷表示的信息存贮在电容器中，在存贮高电平和存贮低电平之间的差别大约为一百万个电子。写操作是把数据放在位线上同时用字线控制这些数据进入电容器而完成；读操作则是靠预充位线和提高字线来实现。单元电容和位线之间的电荷分配，使单元中的信息被破坏，同时使位线电压改变5—30%。读出放大器对两条位线进行比较，以经常检测位线电荷的变化。一条位线上放的是虚单元的电荷，而另一条位线则与被访问的单元相连。虚单元存贮的电荷应与高电平单元和低电平单元的平均电荷一致。

我们集中分析单元版图和电路设计这两个核心问题。对单元版图来说，最重要的是为存贮所需电荷必须的面积和位线间隔，它们决定了读出放大器版图设计的难点。从电路的观点看，我们最关心的是电荷从单元转移到读出放大器的效率和阵列存取连线的 RC 时间常数，位线越长，就有更多的单元与之相连， RC 时间常数和转移效率就越差。这就促使人们把阵列分割成若干小块，但是，这种作法却增加了为连接这些小块所用电路的面积。

我们来研究一下用 $2\mu\text{m}$ 工艺制造的两种不同的试验单元版图，用以对照分析一些不同的问题。第一种单元示于图7.23，为金属字线。由于字线是和晶体管的栅相连，在单元中需要用金属与多晶硅接触。存贮电容器是用MOS电容器做的，大部分电荷存贮在 C_{ox} 中，而一些电荷则存贮在耗尽区（我们的nMOS工艺例子耗尽区很厚，或许对收集因 α 粒子轰击而产生的电荷很有效的）。除非采用耗尽注入，作为存贮电容器一端的极板必须接 V_{DD} ，并且一定要注意保证MOS电容器不存在低于阈值的危险。位线采用扩散区，因此，位线的电阻和电容就都有问题了；增加一个埋层接触则可以用多晶硅而不用扩散区作位线，这种技术已在某些芯片上采用并获得好处。

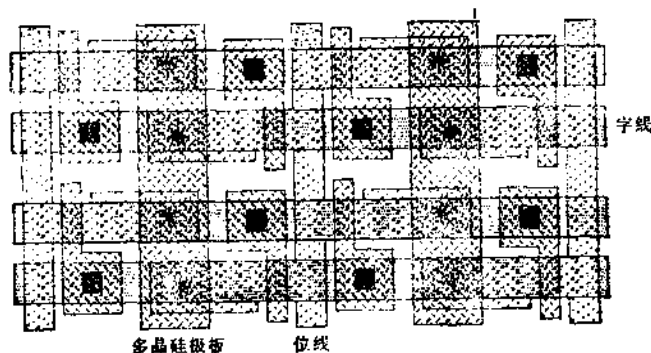


图7.23 金属字线单管RAM版图

图7.24为第二种单元，其电容大致和图7.23中的相同，但不用金属字线而用金属位线。这种版图的面积约为前一种设计的一半，但出现了另外一些问题，例如，很显然，多晶硅字线上信号的传输是相当差的。调节多晶硅的宽度可以减小 RC 效应，而 RC 时间常数对阵列的块数，从而对整个阵列的面积的影响非常大。双层金属对降低 RC 效应也有些帮助，因为这种工艺可以用金属搭接几个单元的多晶硅。但是，第二层金属的间距要比多晶硅的间距或第一层金属的间距大得多，因而这种方法会损失一些面积，并且单

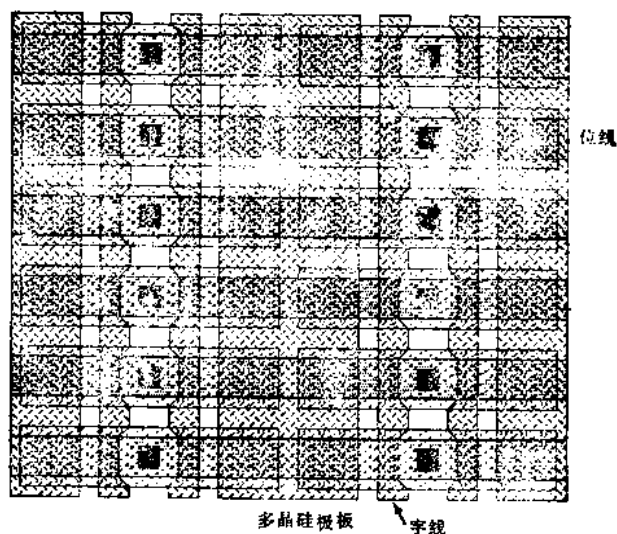


图7.24 金属位线单管RAM的版图

元中也没有多少剩下的面积来放置附加的接触孔。低阻的硅化物对这类电路会有极大的帮助。还应指出, 根据我们的设计规则, 已没有空余的面积来放置 MOS 电容器的耗尽注入了, 所以它的一个极板必须接 V_{DD} , 这就限制了在电容器上存储电压的范围。位线间距已经从 $30\mu\text{m}$ 减小到 $6\mu\text{m}$, 这是另一个重要问题, 它使读出放大器的版图设计极为困难。

商品化的存储器还存在许多别的问题, 例如, 由于阵列要重复如此多次, 设计规则精打细算是合算的, 但是, 这就会导致成品率下降。在单元中金属线对台阶的覆盖性质和数量等这些细节在单元设计中照例也是要考虑的。

图7.25为连在两条位线上的读出放大器, 每条位线各有一个虚单元。在本图中, 位线是逆着阵列方向走的, 如果把位线折叠起来, 它们就互相平行地在阵列的同一边走, 于是某些共模噪声将减小。再进一步以差分方式用两个单元来存储每一位, 就得到非常高的噪声容限, 而耗费的面积则更大, 但至少虚单元已没有必要用了。

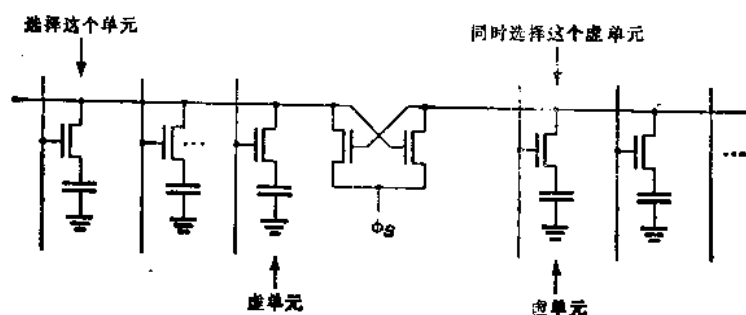


图7.25 单管RAM的位线结构

在图7.25中, 电路的读单元操作是首先使位线预充, 并接通要读出单元的字线和相应的虚单元字线, 然后启动读出放大器, 把电平差放大到全摆幅, 最后把存储的电压写回到访问的单元中。

另一个问题是输出选择器相对于读出放大器的布局问题。即把读出放大器放在列选择器的前面还是后面？读出放大器用的越多，芯片的功耗就越大；反之，读出放大器少则信号电平就低。低的信号电平是可以检测到的，但因为一定要把低电压放大到全逻辑摆幅，增益-带宽乘积限制了总速度。和ROM的情况一样，行存取所需时间比列存取所需时间长得多。

存贮在电容器上的信号电平的大小对RAM的速度和可靠性是关键因素。为了提高存贮电压，有时候使位线自举到高于 V_{DD} 。另一些方法则更多涉及工艺方面。双层多晶硅可使版图紧凑，并用极薄的氧化层作存贮电容器，如图7.26所示。在存贮晶体管下面做高浓度P⁺注入就会使 X_{jmax} 减少而 Q_j 则提高，用这种方法可以得到高电容（High-C）单元。为补偿阈值向上漂移，需要浅的n⁺注入。提高电容器氧化层的介电常数是增加存贮电荷的又一方法，已经试验了Si₃N₄（相对介电常数 $\kappa = 7.5$ ）和Ta₂O₅（ $\kappa = 22$ ）两种方案。

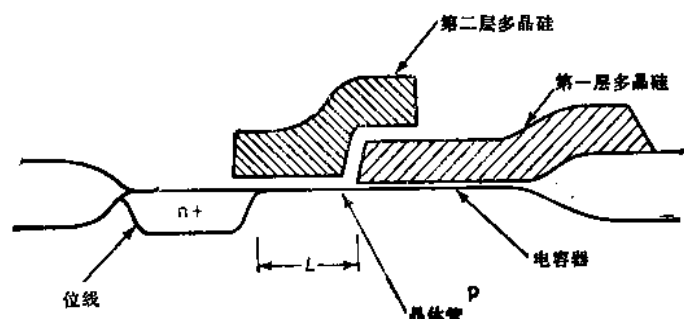


图7.26 用双层多晶硅工艺制造的单管RAM横截面图

商品化的dRAM静态功耗必须为0，为此，在片内产生许多时钟（用时钟链）来启动和停止读放及驱动电路。这些措施在一定意义上也是为了提高速度。

逻辑电路的电路技术往往比商品RAM的电路技术落后数年。本文中两个明显的差异是商品化RAM中更广泛地采用自定时时钟以及对 V_{DD} 更不敏感的读放技术。这就意味着既必须对参考电压采样，又一定要把待测信号变为差分信号。对于静态RAM而言，差分信号是现成的，而在只有单端信号可用的情况下，信号可以与其上一时刻的值进行比较。例如，可以先预充到某个电压（在这种情况下预充到 V_{DD} 是有些好处的），对该电压采样，然后对这采得的电压值和信号建立起来后的电压值进行比较。我们还可以设想，采用这种方法后就避免了在dRAM中使用虚单元。把共用读出放大器两边的位线都预充到约 $V_{DD}/2$ ，然后接通一根字线，被访问的单元就将其位线电平拉成比它初始预充值不是更高就是更低的值。由于这时连在读出放大器另外半边的位线并没有受访问的单元，它的终值和位线启动时的初值一样，所以就可用它作为参考电压。现在两条位线已被驱动成0和 V_{DD} ，重新建立预充电压只要把位线短路就可以了。

7.4 收缩阵列

收缩阵列（systolic arrays）是一种把逻辑电路和存贮器结合起来以产生功能强大的规则结构的设计方法[Kung 79, 79a, 82, Leiserson 81]。收缩阵列是一个小处理器

网络,使通讯和处理器的工作在网络内部协调地组织起来,从而使两者的资源得以充分利用.通讯一般是局部的。“收缩”一词的英文拼写“systolic”来自“systole”,意即收缩.在生物学的意义上,该术语指的是心脏收缩,把血液泵到整个系统.同样,收缩的VLSI系统有节奏地把信息泵向整个阵列.在纯粹的收缩系统¹⁾中,所有处理器之间的通讯都通过一个寄存器.在半收缩系统中,这种收缩的限制有所放宽.收缩阵列的应用场合已有上百种,包括数字滤波、矩阵相乘、矩阵LU因式分解、正则表达式识别、分类排队,以及优先排队等.T.Knng[Fisher 83]关于可编程收缩阵列芯片的设计报告是对这些方面工作的综述.

模式匹配阵列

Foster和Kung[Foster 80]设计了一个模式匹配芯片,说明了许多与收缩子系统设计有关的问题.我们将分析他们的设计,以作为对收缩设计入门的方法.

模式匹配器取两个输入产生一个输出,输入的是样本 P 和字符串 S ,其目标是找出在输入字符串中和全部样本 P 一样的字.这样,当字符串asbksskkkc和样本 $s\Omega k$ (其中 Ω 可代表任何字母)对照比较时就产生输出0001001100.

一种进行模式匹配的算法是使字符串通过样本,并将样本的每一个字母和与之相邻的字符串段进行比较,找出什么时候全部样本和字符串段匹配.这种方法的困难在于必须以全局方式决定是否存在完全的匹配.

较好的一种方法是样本和字符串相碰,如图7.27所示,有关匹配的信息就可以局部地积累而无须全局通讯.由于样本和字符串两者都在移动,必须使它们都各自慢下来,以便新的输入在每逢第二个时钟脉冲时进来,否则它们将会互相错过.这表示在任何一一个时钟脉冲,比较器单元有半数是没有工作的.

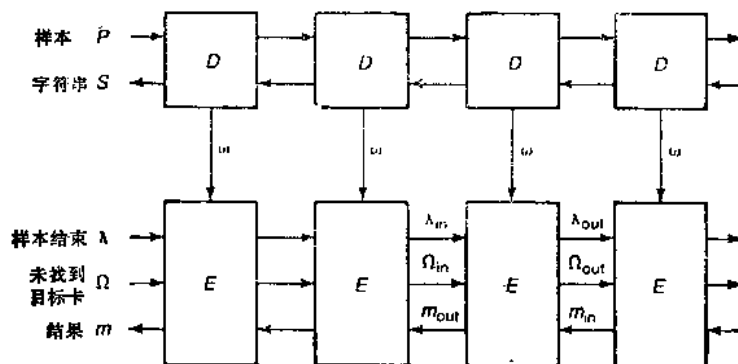


图7.27 收缩模式匹配器的一般结构 [Foster 80]

为了加快产生输出字符串 m ,在样本的最后一个字母后面附上一个符号 λ .从图7.27底部的单元 E ,我们可以看出对每一个时钟脉冲都有

$$\lambda_{in} \rightarrow \lambda_{out} \quad (7.9)$$

$$\Omega_{in} \rightarrow \Omega_{out} \quad (7.10)$$

$$\text{TEMP} \cdot \lambda_{in} + m_{in} \cdot \bar{\lambda}_{in} \rightarrow m_{out} \quad (7.11)$$

1) 流水线的逻辑阵列和细胞或棋盘状的自动系统都是收缩阵列的特例。

阵列优化

已知一个具有收缩能力的逻辑结构, 则有许多种变换可以用来对它进行处理和优化 [Leiserson 83], 下面我们将分析其中的一小部分。对于我们将提出的那些问题, 已经有了用来进行最佳处理的高效率算法。

方法之一是再定时, 这种方法与我们在第六章中遇到过的时钟歪斜变换相似。图 7.29 为一个收缩系统, 其中圆圈代表逻辑块, 矩形为寄存器。如图所示, 我们围绕系统的一部分画一个面。从第六章我们知道可以在子系统的所有输入端加上延迟, 同时从所有输出端减去这个延迟, 这样就得到一个完全等效的外部系统。把这个概念普遍化后, 我们就得到若把寄存器 P 加到每个输入端, 同时从每个输出端除去寄存器 P , 则系统保持等效。这就是再定时。就外部系统而言, 经过再定时后的系统, 其功能和时序是完全一样的。我们以有限冲激响应 (FIR) 数字滤波器为例, 图 7.30 为其示意图。滤波器实现了方程

$$y_t = \sum_{k=0}^{n-1} a_k x_{t-k} \quad (7.16)$$

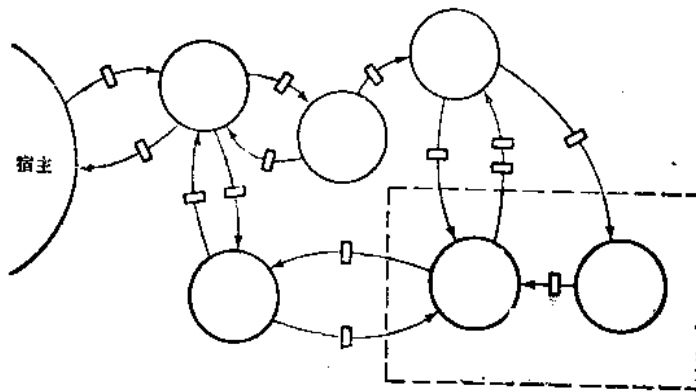


图 7.29 一个收缩系统

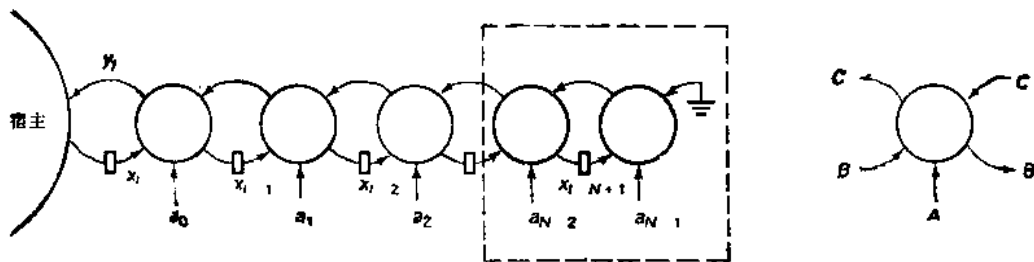


图 7.30 半收缩FIR滤波器

其中 y_t 为 t 时的输出, x_t 为 t 时的输入, a_k 则为滤波器系数。我们用一个“内积”处理器阵列来执行方程 (7.16), 每个处理器完成如下功能:

$$C + AB \rightarrow C' \quad (7.17)$$

以及

$$B \rightarrow B' \quad (7.18)$$

其中 B' 和 C' 为输出, A , B , C 为输入。

图7.30的滤波器是十分简单的,但由于 y_i 必须以行波方式通过 N 个组合逻辑块,性能是很差的。为了清楚起见,在本例中取 $N=5$,并选取最后两个内步进处理器作再定时,如图7.30所示。从输入端移去一个寄存器同时在输出端加上一个寄存器得到图7.31。这样处理后除了主要通道是三个而不是五个内步进处理器以外,功能度是完全一样的。于是,若信号传输经过每一处理器所需的时间为 50ns ,由于再定时,最大时钟速率已由 4MHz 上升到 6.7MHz ;我们基本上无需任何代价就得到几乎高了一倍的最大时钟速率。图7.32为更好的再定时方法,它使最大的主要通道延迟减少到两个模块,性能提高了1.5倍,时钟频率达到 10MHz 。在这种设计方案中取消了不变的输入。这是一种最佳的再定时,但未必是我们所能做的最优方案。

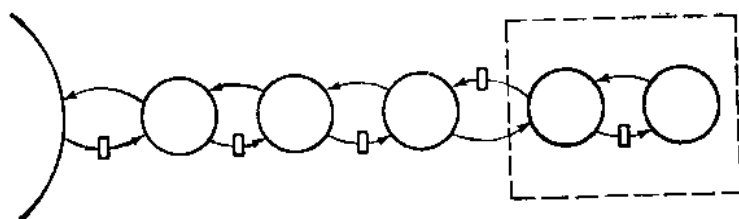


图7.31 经过再定时的FIR滤波器

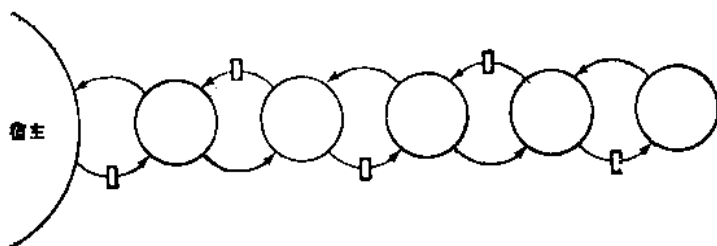


图7.32 优化的再定时FIR滤波器

除了再定时外,还有另外两种方法可以采用,即“减慢”和“阻滞”,这两者都仅改变时钟而不使电路的功能度有所变化。我们首先分析减慢法。

和再定时一样,减慢法是一种普普通通的操作,但却带来潜在的重要结果。为了使一个网络减慢,我们成倍增加寄存器的数量,例如,把原始FIR滤波器电路的寄存器数增加一倍,结果示于图7.33,这是一种流水线的原型。寄存器的数目每增加一倍都会使信号传输到输出端所需时钟脉冲数也增加一倍。虽然主要通道仍保持不变,但注意从奇

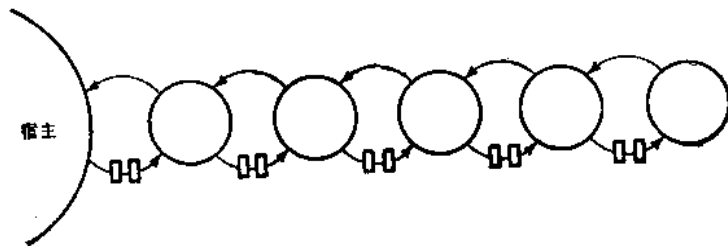


图7.33 经减慢处理的FIR滤波器

数时钟脉冲进入的信号再也不会和偶数时钟脉冲进入的信号相互作用了，这就好像不是一个滤波器，而是两个独立的滤波器了。

网络的信息吞吐量指的是在输出端出现信息的速率，信息由输入端传到输出端所需时间则为等待时间。不幸的是，这些定义有些含糊不清。对于我们研究的滤波器，我们取信号通道的 I/O 速率为信息吞吐量，但这不一定是时钟速率或数据出现在输出端的速率。这是因减慢处理并不改变系统的时钟定时的速率或信息在输出端出现的速率，而是任何单个数据流（通道）可能处理的速率降下来。

如果我们用使寄存器的数量增加 $S - 1$ 倍来减慢系统速度，则通道数就乘了 S 而单个通道的吞吐量则被 S 除。我们以时钟脉冲的整数个数来测量等待时间，对滤波器来说，等待时间就定义为从输入 x_n 到接收到 y_n 的时间。这种定义意味着要处理一些和系统开始工作有关的棘手问题，因为除非我们假设初始条件设置恰当，那么最初几个时钟脉冲可能会产生无效的数据。在大多数信号处理和控制应用中，稳态（一旦流水线已满）响应才是重要的，所以我们在计算等待时间时并没有包括产生无效数据的这段时间。等待时间是取得第一个结果所需要的时间（假设初始条件已正确设置），吞吐量则为其后的结果到达速率（和一个数据流输入相应）。在图 7.30 中，吞吐量和时钟率都是 4MHz；在图 7.33 中，时钟率是 4MHz，而对两个相同通道中的每一个，吞吐量都是 2MHz；图 7.30 的 $N = 5$ ，等待时间为 250ns，而图 7.33 的等待时间则为 500ns；对图 7.32 所示电路来说，吞吐量和时钟率均为 10MHz，而等待时间是 100ns（这时间可以是 50ns，但我们要求等待时间以时钟周期为单位来衡量，用 10MHz 时钟来同步电路一次就是 100ns，因此等待时间定为 100ns）。

在图 7.33 中应用减慢处理后，还可以对电路进行再定时，图 7.34 所示为优化再定时。主要通道已缩减为通过一个处理器的传输时间，这就使前面得到的最好结果又提高了一倍，并比最初的性能好五倍，这里，时钟率为 20MHz，两个通道中的每一个吞吐量都是 10MHz，以及等待时间也是 100ns。

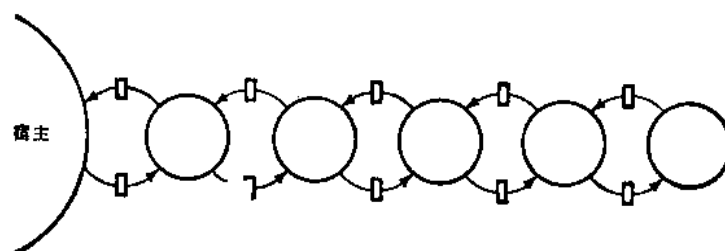


图 7.34 经优化减慢和再定时处理的 FIR 滤波器

从研究内积处理器的内部工作可以获得又一种设计方案。从 B 产生 B' 只要一条连线，而产生 C' 则要一条相加线和一条相乘线，因此，如果初始方案中有一个信号必须是全局的，这信号应是 x 而不应是 y 。在这种设计中，我们将再次从再定时得到好处。一种具有一条全局连线的再定时电路示于图 7.35。

一条连线只是一条连线而已，它本身并不具有使信号一定要沿着某个特殊方向流动的特性，在图 7.36 中 x 是从右流到左而不是从左流到右（树或其它拓扑已能用于分配 x ），这是又一种等效系统。为使这种系统流水线工作，要求第三种操作，这就是阻滞

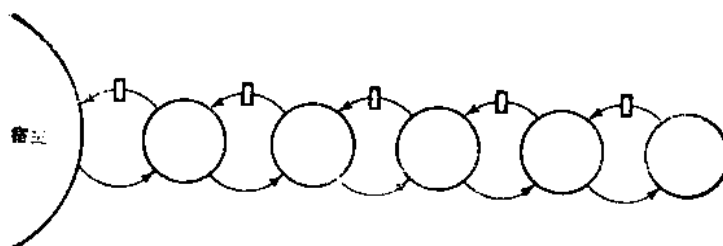


图7.35 另一种再定时FIR滤波器

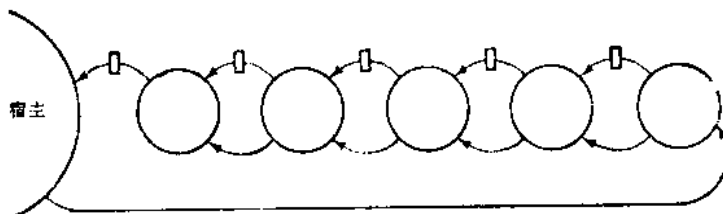


图7.36 改变滤波器的拓扑结构

操作。阻滞操作是把寄存器放在系统的输入端或输出端以增加等待时间，图7.37即为示意图。由再定时得到一种如图7.38所示的流水系统，在这种电路中时钟速率和吞吐量都已达到20MHz，为其最佳值。为了使吞吐量有这么大的提高，等待时间增加到500ns。从图7.30变到图7.32，由于最原始的系统已经是次优的，我们有得无失。而在优化电路图7.32、图7.34和图7.38中，却要在通道的数量、每一通道的吞吐量，以及等待时间之间进行折衷。表7.1综述了几个关键的结论。

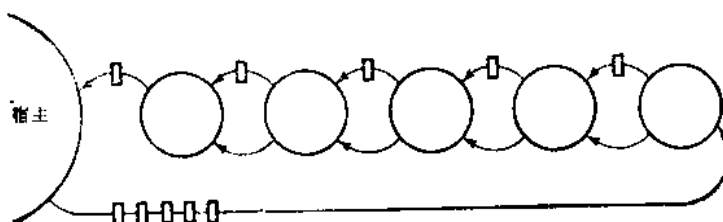


图7.37 使滤波器阻滞

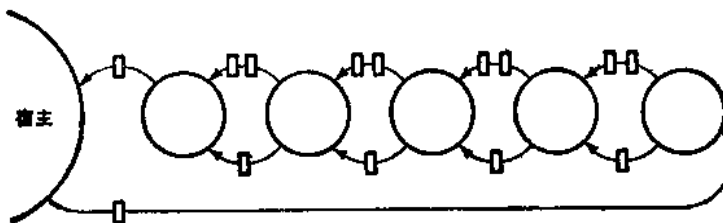


图7.38 用阻滞加再定时得到最大吞吐量

其它一些如逻辑重制、资源共享等等诸如此类的变换扩大了上述三种变换。Charles Leiserson[Leiserson 83]提出了一种引起人们兴趣的方法来利用这些技术，他建议设计一个具有特定性能的复杂系统的方法，即，首先设计一个从直观判断就是有吸引力的和

表7.1 FIR滤波器的特性

图 号	时钟速率 (MHz)	吞吐量 (MHz)	等待时间 (ns)	通道数
7.30	4	4	250	1
7.32	10	10	100	1
7.33	4	2	500	2
7.34	20	10	100	2
7.38	20	20	500	1

简单的但确保一定能工作的系统，然后用已经证明是有效的变换进行处理，把设计优化为一个任务要求的系统。

7.5 小 结

设计大型阵列涉及许多不同层次的概念，我们已经研究的技术包括：

- X单元版图设计，
- 虚接地，
- 离子注入编码，
- PLA逻辑最小化，
- PLA和ROM的折叠，
- 输入编码，
- 收缩系统的再定时、减慢，和阻滞。

大型阵列的优化设计是极为重要的，因为这种阵列常常由于面积和速度两者的改善而使VLSI系统得以实现。在存储器阵列设计方面，单元尺寸的优化可以增加外围电路

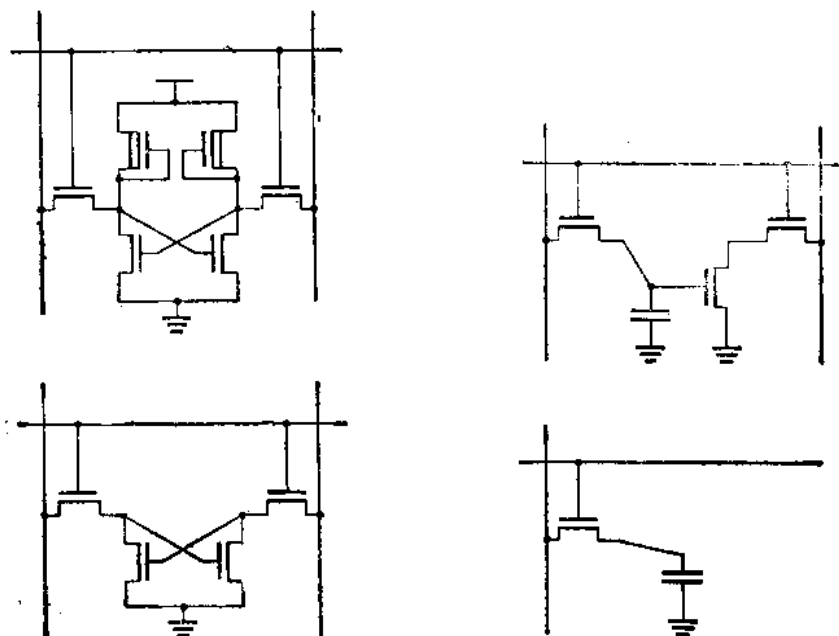


图7.39 奇妙的RAM单元演变图。下一步将如何发展

的复杂程度和使用的面积。RAM单元从六管静态到单管动态的演变示于图 7.39, 该图尽管是事后归纳出来的, 但仍引起人们的兴趣, 下一步发展是什么尚不得而知。

习 题

- 7.1 设计一个电路, 使其输入有跳变时输出将上升。该电路的工作必须与输入波形的变化速度无关, 并有一个复位输入。
- 7.2 利用图7.40所示电路图, 为例2.9的ROM设计一个读出放大器。△

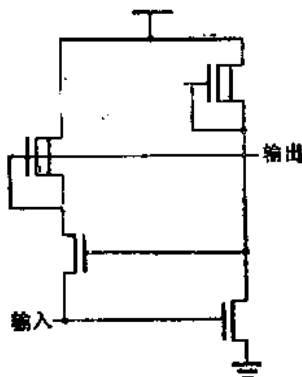


图 7.40

- 7.3 预充nMOS电路常有的问题是高电平的确切值与预充开始了多长时间有关。次开启漏电是一种真正的故障, 在作模拟时也会显示出这个问题。试讨论这一现象在哪些方面会引起困难。△
- 7.4 实验测定存储电路对噪声灵敏度的一种方法是使电源快速变化, 例如从4V变到6V, 一般情况下就可找出设计中那些抗噪声强度不怎么够的部分。这种方法不能发现的是哪些类型的问题?△△
- 7.5 用PLA设计一种二进制到BCD的译码器, 并用折叠和逻辑最小化技术来减小阵列的面积。
- 7.6 按内容寻址存储器(CAM)从概念上说是一种工作方式正好和正常RAM相反的阵列。正常RAM在已知地址的条件下输出单元的内容, 而CAM则是输出和已知内容相符的地址。CAM的关键部分是匹配单元(match cell)。图7.41就是一种动态CAM单元。试解释这CAM单元是怎样进行读和写的。

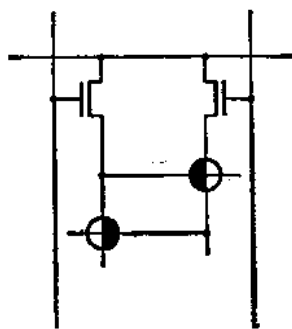


图 7.41

- 7.7 对7.4节中讨论的两种模式匹配单元中的每一种, 画出晶体管级CMOS电路。
- 7.8 用最少数目的乘积项重新设计图7.10输入编码PLA的逻辑, 分析行和列折叠的可能性。
- 7.9 再定时技术可以用来从一种概念上看简单但性能不佳的设计得到有意义的高性能设计。假定有一个最简单的 $N \times N$ 平行乘法器, 它执行 $N - 1$ 重复移位并与被乘数相加, 每次把进位送过整

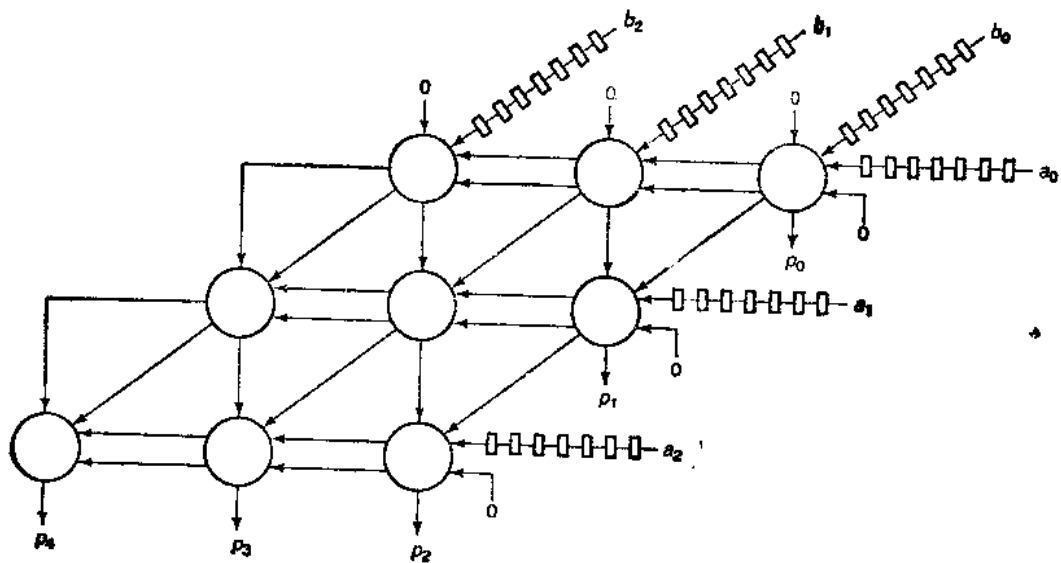


图 7.42

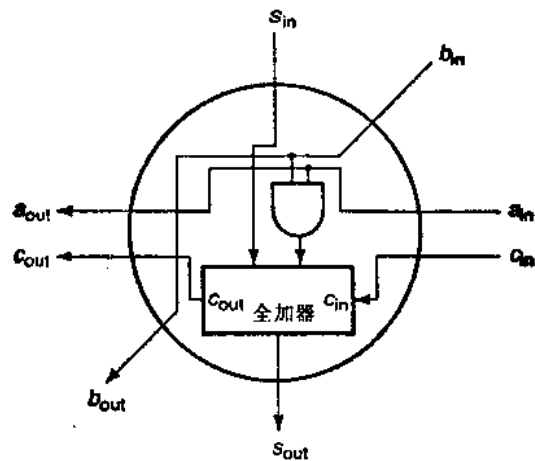


图 7.43

个字长。对于 $N = 3$ 的这种乘法器电路示于图 7.42，使用的单元为图 7.43。在每一个输入端放了七级寄存器来实现流水作业。试对图 7.42 的乘法器进行再定时处理以获得一个接受 3 位乘数和 3 位被乘数，并且在每一时钟脉冲产生 5 位乘积的流水线并行乘法器，最后使该乘法器的吞吐量达到最大。

7.10 一种五极点无限冲激响应(IIR)滤波器具有转移函数

$$H(z) = \frac{1}{1 - \sum_{i=1}^5 a_i z^{-i}}$$

利用解下面的递归式，在时间域中实现这种滤波器：

$$y_n = a_1 y_{n-1} + a_2 y_{n-2} + \dots + a_5 y_{n-5} + x_n$$

其中 x_n 和 y_n 分别为滤波器的输入和输出。递归用图 7.44 的电路实现，该电路的反馈回路中用了



图 7.44

五点FIR滤波器，反馈电路的输入和输出则为 y_n 和 w_n ，其中

$$w_n = \sum_{i=1}^5 a_i y_{n-i}$$

图7.32和图7.38为实现该反馈电路的两种可能方案。若设寄存器的延迟可以忽略不计，并且所有单元(包括图7.44的加法器)的传播延迟时间为50ns，请选择一种电路，使FIR滤波器的吞吐量最大，计算该最大吞吐量，并简单说明等待时间和FIR反馈电路吞吐量对FIR滤波器吞吐量的限制。 Δ

参考文献

- [Anderson 82] J.M.Anderson, B.L. Troutman, and R.A. Allen, "A CMOS LSI 16×16 Multiplier/Multiplier-Accumulator," *IEEE International Solid-State Circuits Conf.*, 124—125, San Francisco, Calif., 1982.
- [Ayres 83] R.F.Ayres, *VLSI Silicon Compilation and the Art of Automatic Microchip Design*, Prentice-Hall, Englewood Cliffs, N.J., 1983.
- [Cook 79] P.W. Cook, S. E. Schuster, J.T.Parrish, V. DiLeonardo, and D.R.Freedman, "1 μ MOSFET VLSI Technology, Part III," *IEEE Trans. Electron Devices* ED-26, 333—345, April 1979.
- [Cook 79b] P.W.Cook, C.W.Ho, and S.E.Schuster, "A Study in the Use of PLA-Based Macros," *IEEE J. Solid-State Circuits* SC-14: 833—840, October 1979.
- [Dennard 68] R.H.Dennard, "Field-Effect Transistor Memory," U. S. Patent 3,387,286, June 4, 1968.
- [Fisher 83] A.L.Fisher, H.T.Kung, L.M.Monier, H.Walker, and Y.Dohi, "Design of the PSC, A Programmable Systolic Chip," *Proc. Third Caltech Conf. on VLSI*, 287—302, Pasadena, Calif., 1983.
- [Fleisher 75] H.Fleisher and L.J.Maissel, "An Introduction to Array Logic," *IBM J. Res. Develop.* 19, 98—109, 1975.
- [Foster 80] M.J.Foster and H.T.Kung, "The Design of Special-Purpose VLSI Chips," *IEEE Computer* 13, 26—40, 1980.
- [Guttag 82] K.M.Guttag, J.F.Sexton, K.S.Chang, "A 16b Microprocessor with a 152b Wide Microcontrol Word," *IEEE International Solid-State Circuits Conf.*, 120—121, San Francisco, Calif., 1982. (TMS 99000)
- [Hachtel 82] G.D.Hachtel, A.R. Newton, and A. L. Sangiovanni-Vincentelli, "An Algorithm for Optimal PLA Folding," *IEEE Trans. Computer-Aided Design* CAD-1, 63—77, 1982.
- [Hardee 81] K.C.Hardee and R.Sud, "A Fault-Tolerant 30 ns/375mW16K $\times$ 1 NMOS Static RAM," *IEEE J. Solid-State Circuits* SC-16, 435—443, 1981. (A number of amazing circuits)
- [Hardee 84] K.Hardee, M. Griffus, and R.Galvus, "A 30 ns 64 K CMOS RAM," *IEEE International Solid-State Circuits Conf.*, 216—217, San Francisco, Calif., 1984.

- [Hong 74] S. J. Hong, R. G. Cain, and D. L. Ostapko, "MINI: A Heuristic Approach for Logic Minimization," *IBM J. Res. Develop.* 19, 143-158, 1974.
- [Hudson 82] E. L. Hudson and S. L. Smith, "An ECL Compatible 4K CMOS RAM," *IEEE International Solid-State Circuits Conf.*, 248-249, San Francisco, Calif., 1982. (This paper contains some "typos" concerning the types of bipolars used, so be careful.)
- [Isobe 81] M. Isobe, Y. Uchida, K. Maeguchi, T. Mochizuki, M. Kimura, H. Hatano, Y. Mizutani, and H. Tango, "An 18 ns CMOS/SOS 4K Static RAM," *IEEE J. Solid-State Circuits* SC-16, 460-465, 1981.
- [Itoh 84] K. Itoh, R. Hori, I. Etoh, S. Asai, N. Hashimoto, K. Yagi, and H. Sunami, "An Experimental 1Mb DRAM with On-Chip Voltage Limiter," *IEEE International Solid-State Circuits Conf.*, 282-283, San Francisco, Calif., 1984.
- [Jaing 81] C.-L. Jaing and R. Plachno, "A 32K Static RAM Utilizing a Three-Transistor Cell," *IEEE International Solid-State Circuits Conf.*, 86-87, New York, N.Y., 1981.
- [Kang 81] S. Kang, "Synthesis and Optimization of Programmable Logic Arrays," Technical report no. 216, Computer Systems Laboratory, Stanford University, 1981.
- [Kawagoe 76] H. Kawagoe and N. Tsuji, "Minimum Size ROM Structure Compatible with Silicon-Gate E/D MOS LSI," *IEEE J. Solid-State Circuits* SC-11:360-364, 1976.
- [Kertis 84] R. A. Kertis, K. J. Fitzpatrick, and Y.-P. Han, "A 59ns 256K DRAM using LD² Technology and Double Level Metal," *IEEE International Solid-State-Circuits Conf.*, 96-97, San Francisco, Calif., 1984.
- [Kung 79] H. T. Kung and C. E. Leiserson, "Systolic arrays (for VLSI)," In I. S. Du and G. W. Stewart (eds.), *Sparse Matrix Proceedings* 1978, Society for Industrial and Applied Mathematics, pp. 256-282, 1979. (An earlier version appears in Chapter 8 of Mead and Conway under the title, "Algorithms for VLSI Processor Arrays.")
- [Kung 79a] H. T. Kung, "Let's Design Algorithms for VLSI Systems," *Proceedings of the Caltech Conference on Very Large Scale Integration*, Charles L. Seitz (ed.), Pasadena, Calif., 55-90, 1979.
- [Kung 82] H. T. Kung, "Why Systolic Architectures," *IEEE Computer Magazine*, 37-46, 1982.
- [Leiserson 81] C. E. Leiserson, *Area-Efficient VLSI Computation*, Ph.D. dissertation, Department of Computer Science, Carnegie-Mellon University, October 1981. Published in book form as part of the ACM Doctoral Dissertation Award Series by the MIT Press, Cambridge, Mass., 1983.
- [Leiserson 83] C. E. Leiserson and J. B. Saxe, "Optimizing Synchronous Systems," *J. VLSI and Computer Systems* 1, 41-67, 1983.
- [Lewyn 84] L. L. Lewyn and J. D. Meindl, "Physical Limits of VLSI DRAMs," *IEEE International Solid-State Circuits Conf.*, 160-161, San Francisco, Calif., 1984.
- [Lopez 80] A. D. Lopez and H. F. Law, "A Dense Gate Matrix Layout Style for MOS LSI," *IEEE International Solid-State Circuits Conf.*, 212-213, San Francisco, Calif., 1980.
- [Lund 70] D. Lund, C. Allen, S. Anderson, and G. Tu, "Design of a Megabit Semiconductor Memory System," *AFIPS Fall Joint Computer Conf.*, 53-62, 1970. (4T RAM)
- [Lyon 80] R. F. Lyon, private communication, 1980.
- [Mashiko 84] K. Mashiko, T. Kobayashi, W. Wakamiya, M. Hatanaka, and M. Yamada, "A 70ns 256K DRAM with Bitline Shielding Structure," *IEEE International Solid-State Circuits Conf.*, 98-99, San Francisco, Calif., 1984.
- [Masuoka 84] F. Masuoka, S. Arizumi, T. Iwase, M. Ono, and N. Endo, "An 80ns 1Mb ROM," *IEEE International Solid-State Circuits Conf.*, 146-147, San Francisco, Calif., 1984.
- [May 76] P. May and F. C. Schierick, "High-Speed Static Programmable Logic Array in LOCMOS," *IEEE J. Solid-State Circuits* SC-11, 365-369, 1976.
- [Minato 81] O. Minato, T. Masuhara, T. Sasaki, Y. Sakai, and K. Yoshizaki, "A High-Speed Hi-CMOS 4K Static RAM," *IEEE J. Solid-State Circuits* SC-16, 449-453, 1981.
- [Minato 84] O. Minato, T. Masuhara, T. Sasaki, Y. Sakai, and T. Hayashida, "A 20ns 64 K CMOS SRAM," *IEEE International Solid-State Circuits Conf.*, 222-223, San Francisco, Calif., 1984.
- [O'Connell 77] T. K. O'Connell, J. M. Hartman, E. D. Errett, G. S. Leach, and W. C. Dunn, "A 4

- K Static Clocked and Nonclocked RAM Design," *IEEE International Solid-State Circuits Conf.*, 14—15, Philadelphia, Pa., 1977.
- [Ochii 77] K. Ochii, Y. Suzuki, M. Ueno, K. Sato, and K. Asahi, "C²MOS 4K Static RAM," *IEEE International Solid-State Circuits Conf.*, 18—19, Philadelphia, Pa., 1977.
- [Patil 79] S. S. Patil and T. A. Welch, "A Programmable Logic Approach for VLSI," *IEEE Transactions on Computers* C-28, 584—601, 1979. (SLAs)
- [Remshardt 76] R. Remshardt and U. G. Baitinger, "A High Performance Low Power 2048-Bit Memory Chip in MOSFET Technology and Its Application," *IEEE J. Solid-State Circuits* SC-11, 352—359, 1976.
- [Regitz 70] W. M. Regitz and J. Karp, "Thres Transistor Cell, 1024 bit, 500ns MOS RAM," *IEEE J. Solid-State Circuits*, pp. 42—43, 1970.
- [Schmidt 65] J. S. Schmidt, "Integrated MOS Random-Access Memory," *Solid-State Design*, pp. 21—25, 1965.
- [Schnookler 80] M. S. Schnookler, "Design of Large ALUs Using Multiple PLA Macros," *IBM J. Res. Develop.* 24, 2—14, January 1980.
- [Sud 81] R. Sud and K. C. Hardee, "Designing Static RAMs for Yield as Well as Speed," *Electronics* 54, 121—128, July 28, 1981.
- [Taguchi 84] M. Taguchi, S. Ando, S. Hijiya, T. Nakamura, S. Economo, and T. Yabu, "A Capacitance-Coupled Bit-Line Cell for Mb Level DRAMs," *IEEE International Solid-State Circuits Conf.*, 100—101, San Francisco, Calif., 1984.
- [Terman 71] L. M. Terman, "MOSFET Memory Circuits," *Proc. IEEE* 59, 1044—1058, 1971.
- [Wada 78] T. Wada, O. Kudoh, Y. Nagahashi, and S. Matsue, "A 15-ns 1024-Bit Fully Static MOS RAM," *IEEE J. Solid-State Circuits* SC-13, 635—639, 1978.
- [Weinberger 67] A. Weinberger, "Large-Scale Integration of MOS Complex Logic: A Layout Method," *IEEE J. Solid-State Circuits* SC-2, 182—190, 1967.
- [Wong 81] J. Wong, M. Ebel, and P. Siu, "A 45-ns Fully Static 16K MOS ROM," *IEEE J. Solid-State Circuits* SC-16, 592—594, 1981.
- [Wooley 80] B. A. Wooley, "Course notes for EECS 243," University of California, Berkeley, unpublished, 1980.
- [Yamada 84] J. Yamada, T. Mano, J. Inoue, S. Nakajima, and T. Matsuda, "A Submicron VLSI Memory with a 4b-at-a-Time Built-in ECC Circuit," *IEEE International Solid-State Circuits Conf.*, 104—105, San Francisco, Calif., 1984.
- [Zippel 83] R. E. Zippel, "A Survey of Static Memory Techniques," unpublished, 1983.

第八章 超大规模集成系统的微结构

微结构所涉及的内容介于宏结构（用户可见的最低层硬件）和工艺技术（MOS VLSI）之间。本章介绍用 MOS VLSI 实现的数字系统设计中的某些重要问题。系统结构和许多课题有关，其中每个课题都值得像以前几章中处理 MOS 电路那样详细展开，但本章还是主要进行综述和举例，同时也列出有关数字系统结构，逻辑设计以及微程序设计等方面的文献目录，以促使读者研究这些有关的原始资料。

结构原理是电路设计师理所当然的研究范围，这正如电路原理对学习计算机结构显然是合适的一样。Carver Mead 和 Lynn Conway 在他们的著作“Introduction to VLSI Systems”是这样描述 VLSI 芯片的设计过程的（Mead 80, p.89）：“设计是由相继的后一层次设计师以“自上而下”且充分理解的方式完成的……。例如，进行集成系统的“逻辑设计”，也许最好是设想为寻找一些技术和发明，这些技术和发明使集成器件和电路的物理、拓扑，以及几何性质与数字 VLSI 系统的要求的特性最紧密地联接在一起。”

这不是一种 VLSI 系统设计的严密观点。成功地实现 VLSI 系统，有赖于对连线、晶体管、逻辑，以及系统结构等之间的协调。因此，对系统设计采用综合的方法才是最佳方案。

我们分割和思考系统的方法是存在矛盾的。当设计工程师必须用强有力的智力抽象过程把问题分割为小到足以方便地领会的程度时，任何时候都不能忘记抽象化只是在头脑中人为的分割。实际上，我们是有意地约束和限制设计方式，从而可以形成正确的抽象，这就是方法的全部含义。抽象是有目的的，但其本身只是实现目的的一种手段。毫无疑问，我们的智力障碍未必反映自然的全部一致性，于是抽象就既是可做到的又是受限制的。在许多情况下，最深刻的理解和发明是超越我们已建立的框框——在这个过程中常常创立新的规矩。在本书的前七章中，我们研究了许多的抽象层次，然而，如果仔细地分析这些层次，就会发现每一个层次都是与其它层次相互联系的。

在经济上和结构上，VLSI 都给设计师们提供了新的灵活性，可惜还没有人理解如何充分利用这些灵活性。即使对于由一个数据通道和一个控制通道组成的普通计算机，不同结构的好坏也是大有争论的。下面我们将利用常规计算机竞争中出现的技术作为进一步研究 VLSI 微结构的基础。

图 8.1 为标准的微处理器，它由数据通道，控制通道和 I/O 组成。由于数据通道是最简单的部件，我们将由此开始分析。

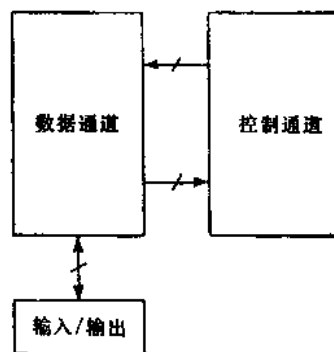


图 8.1 包含有一个数据通道和一个控制通道的简单微处理器

8.1 数据通道设计

在本章所有的分析中，都假定了我们首要的设计目标是用最小的芯片资源得到最大的计算能力。这里所说的芯片资源主要是指芯片面积和功耗（设计资源的限制也是要考虑的）。性能和资源构成设计空间的两个轴。对于一组确定的资源存在一个可达到的最佳性能，所以，芯片资源的函数在设计空间构成了一条可达到的最佳性能曲线。对于一个特定的设计，资源或性能的约束确定了设计的目标点。

在数据通道设计中，主要的微结构设计变量是性能、各个逻辑元件的类型、每种功能类型的元件数量以及元件之间的通信通道。以下一些例子为 ALU、移位器和寄存器堆等功能元件。ALU 的参量包含有该 ALU 可执行的功能类型，它的宽度（指位串行，如 8 位、16 位等等）和它的操作速度。对于移位器，其主要参数是它的宽度、能左移的位数和能右移的位数。在寄存器堆中，我们所考虑的则是它的宽度、寄存器的数量、读出端口的数目、写入端口的数目和读/写端口的数目。

算术逻辑单元

实际上所有微处理器都有一个中央算术逻辑单元，以执行基本计算所必须的逻辑运算和算术运算功能。此外，许多微处理器还包含有少量的通用元件以执行更为特殊的功能，如增量器，相加器或比较器。

一个通用 ALU 往往含有三个基本单元：逻辑功能发生器，进位链和末级异或门。输入到 ALU 的数据通常是两个在位置上配对的寄存器内容。逻辑功能发生器的作用是向进位链和异或门提供输入信号。这些输入信号代表了两个输入变量的各种功能。两个输入变量有 16 种可能的逻辑功能。为了在 16 种功能中进行选择需要 4 条控制线。图 8.2 所示为一个普通的功能发生器。它的两个输出信号 P 和 G 分别代表算术操作时送到进位链的进位传播输入和进位产生输入，同时利用四条控制线可以控制并实现其它的逻辑功能。

可以把功能块看成更一般的查表技术的一个特例。例如，要求把表做成对数运算。常数发生器也是查表的一种特殊情况。在这种情况下，在 PLA 或 ROM 中存储重要数值（如 1，0，-1）和特殊地址，以便通过控制通道方便地寻址。

一个位译码器具有相似的作用。然而，它较为紧凑但通用性较小。对于一个 32 位数据通道来说，一个位译码器接受 5 个地址线来选择 32 条数据通道中的某一条。用优先权编码器可以实现相反的功能，这种编码器对一个指定的位检测其数据通道，并利用控制通道对其位置进行编码。这些功能单元的作用在于减少控制通道和数据通道之间的连线数目。

在许多数据通道电路中都有“滚筒式移位器”（barrel shifter）。图 8.3 示出了一种简单的滚筒式移位器。滚筒式移位器可使数据通道字转动任意位数。

在一个 ALU 中，进位链往往是最使速度受限制的单元，这是因为数据通过它时必须横向跨过 ALU 的宽度。在标准 MOS 逻辑电路中，采用了带有传输器件的曼彻斯特结构。如果要将 A 和 B 两个数相加，对于每一位，都有进位输入和进位输出两种可能性。

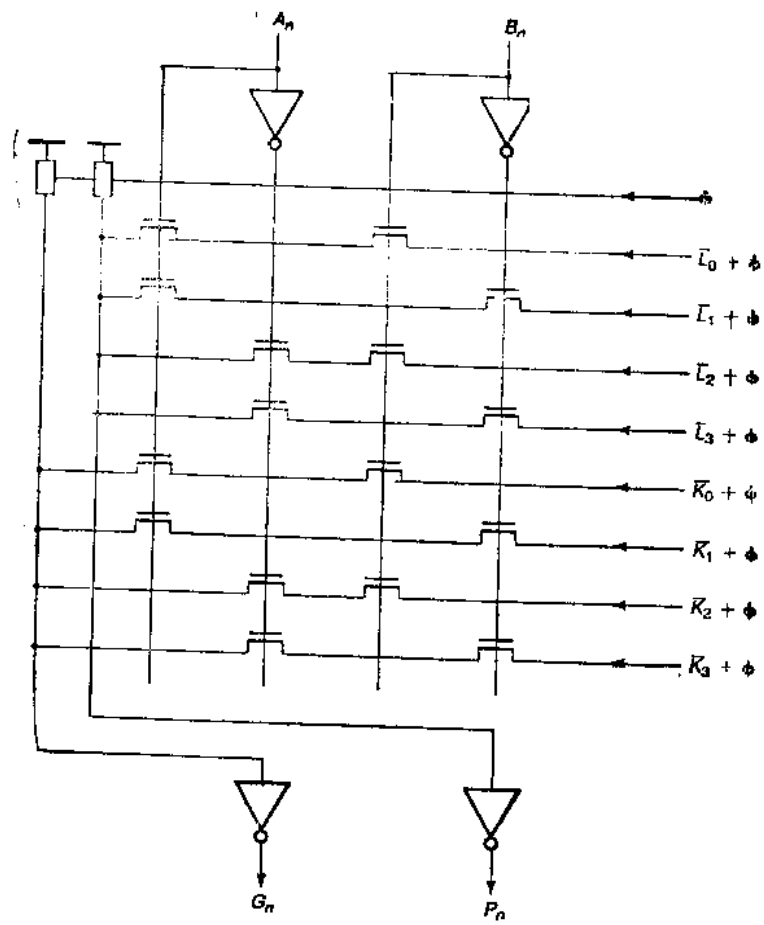


图8.2 一个逻辑单元的功能发生器对

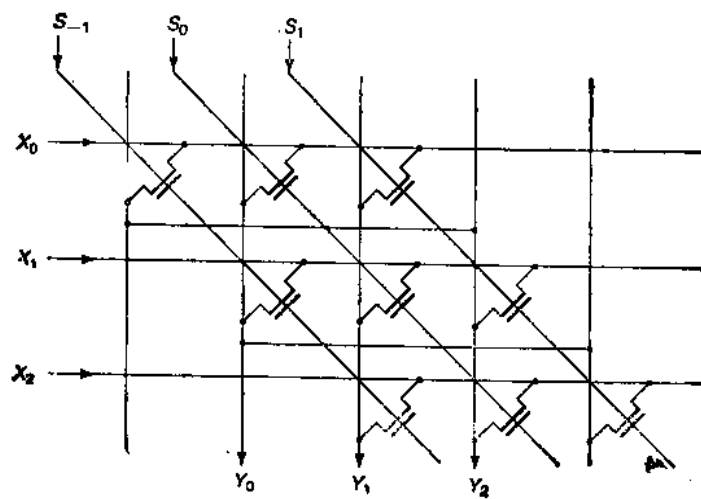


图8.3 滚筒式移位器

如果 A_i 和 B_i 为“1”，则不论进位输入如何，均产生进位输出。在这种情况下，有可能产生进位输出而不需要等待进位输入信号。如果 A_i 和 B_i 均为“0”，则进位输出为“0”，这时，还是和进位输入无关。在这种情况下，立即去掉进位输出信号较为保险。如 $A \oplus B$ 为“1”，只有在进位输入为“1”时，进位输出才是“1”。这就是说，进位被“传播”了。表 8.1 总结了这些条件。

我们有

$$K_i = \bar{A}_i \cdot \bar{B}_i \tag{8.1}$$

$$G_i = A_i \cdot B_i \tag{8.2}$$

和

$$P_i = A_i \oplus B_i \tag{8.3}$$

它的和 S_i 由下式给出：

$$S_i = P_i \oplus C_{i-1} \tag{8.4}$$

其中 C_{i-1} 是进位输入。

图 8.4 给出了采用预充的曼彻斯特进位链的典型电路。为了防止串联传输器件的二次延迟作用，按照一定间隔插入了缓冲器。图 8.5 给出了一种缓冲器电路，它避免了与临界通路串联时附加的缓冲器延迟，而且可实现要求的电平恢复功能。利用这种方法插入缓冲器（每 2—4 级插入一个）对取得最佳性能是有利的。

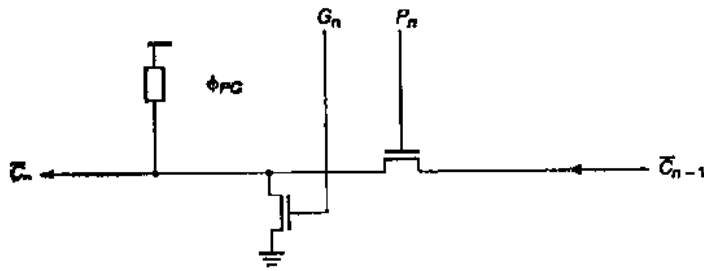


图 8.4 采用预充的曼彻斯特进位链的一级

表 8.1 曼彻斯特进位条件		
A_i	B_i	操 作
0	0	封锁进位
0	1	传播进位
1	0	传播进位
1	1	产生进位

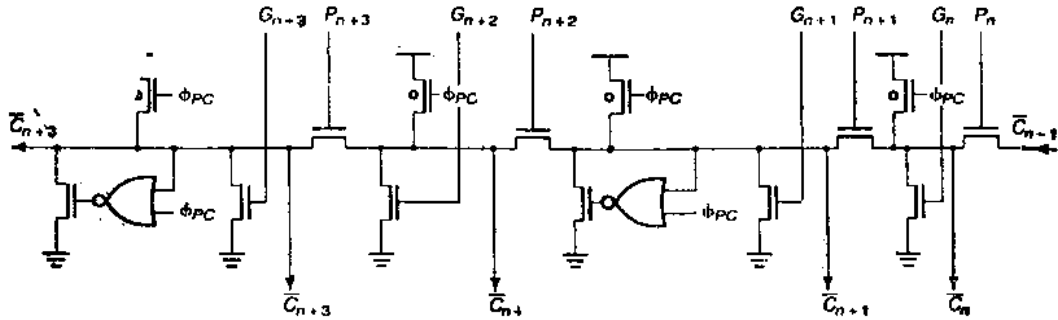


图 8.5 在非主要通路中的进位链信号恢复缓冲器

先行进位技术对加速进位传播是有用的。这种技术是在同一时刻对许多位进行了进位计算而不是只算一位。在图 8.6 中，给出了一种利用曼彻斯特结构的性能实现先行进

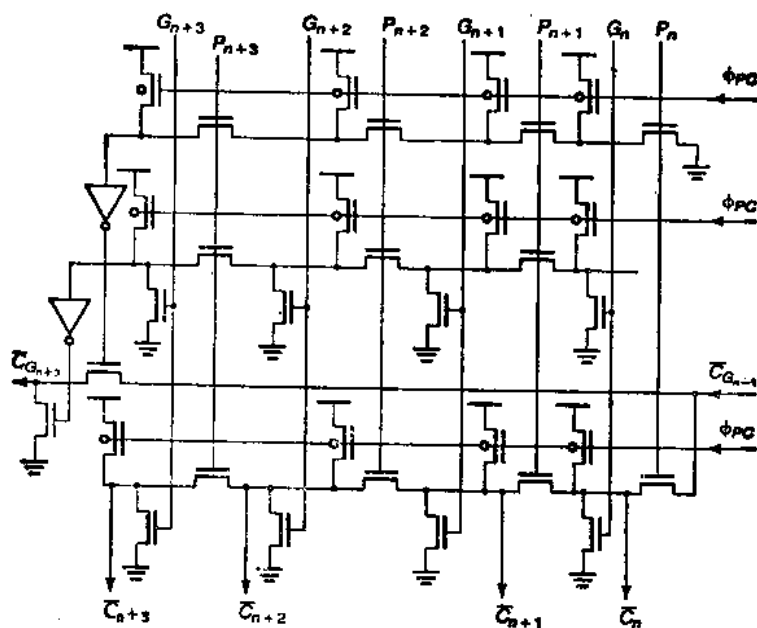


图 8.6 采用曼彻斯特结构的先行进位链组

位的 MOS 电路。对于特定用途，可以优化先行进位的分组规模。然而，经验表明，四位作为一组往往是最优的，特别是在先行进位产生通路中不用进位缓冲器时更是如此。

条件选择器是用于控制电路和数据电路的一种十分通用技术。如果当二进制输入为高时要得到一种操作，而当输入为低时要得到另一种操作，则有可能立即做到这两点（在分离的硬件中）。当输入一出现就选出正确的答案。共用某些硬件来计算这两种可能性通常是可能的。对上述做法存在一些限制，并且，它随输入位数的增加而极为难以控制，然而在很多情况下，适合设计为 VLSI。选择器是经济的，硬件复制是直接了当的。图 8.7 给出了采用这种技术的加法器，其中假定中间进位是双向的，并选出正确答案。

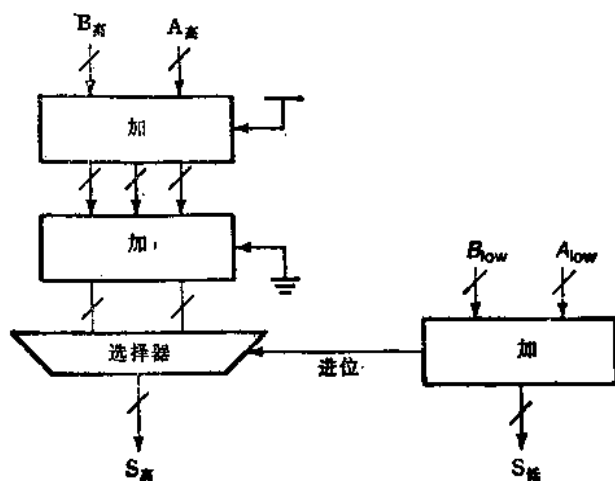


图 8.7 进位选择加法器

末级异或门往往给电路设计师带来困难。这是因为在 P 和 G 开始计值时刻和最后的局部进位输入时刻之间，输出会有假信号，所以一般不大可能进行预充，这与函数发生器和进位链不大一样。如果有合适的时钟边沿可用，则在建立输入信号之后就可选通异或门。然而，假定ALU一定是径向的（通常情况下是这样的），设计师就必须在静态电路中进行选择。图8.8给出了一种采用反相器和传输器件的静态电路。通常 P 端的反相器常安排恰当，所以关键通路只包括局部进位反相器。

ALU的表现宽度常常受限于系统的上层结构。然而，往往可以用一个实体ALU仿真这个“虚拟”ALU，该实体ALU比较小一些，但是要求若干通路以实现与虚拟器件同样的操作。例如，在早期的4位微处理器中，曾采用位串行ALU。在某些16位微处理器单元中使用了8位ALU。采用这种技术的最好例子是手动计算器，在那里一定是使用位串行运算器以便实现精确的计算。芯片面积和功耗大体上正比于ALU的宽度。这就说明了窄的ALU较为有利。然而，通过一个宽ALU的延时，只是按ALU宽度的对数关系变化。这就说明，单通路通过宽ALU要比多通路通过窄ALU更为有利。

对于数据通道中某些特殊功能来说，整个ALU功能的子集往往是有用的。寄存器增量器就是这种功能的一个实例。图8.9给出了一种增量寄存器电路，这种寄存器除了在传输器件组以外并不要求行波进位。通过这种传输器件组的进位传播，可以用图8.5中所示的进位缓冲器来提高速度。

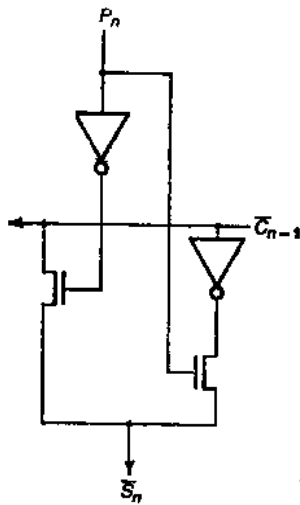


图8.8 在ALU中末级异或门的静态电路

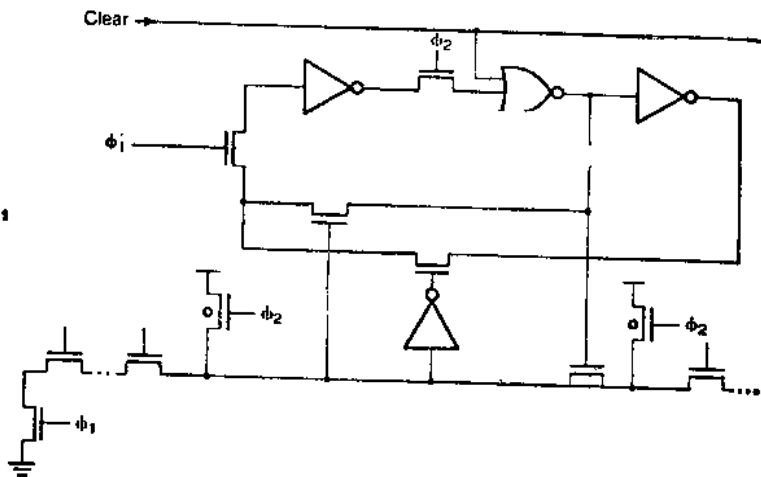


图8.9 与图1.58中电路相似的预充增量单元

寄存器堆

微处理器中的寄存器堆与第七章中所讨论的RAM相似，然而，约束条件往往极不相同。例如，由于位线间距受到ALU的限制，致使RAM单元的版图只能是次优的。

寄存器堆设计中主要考虑的是所要求的端口数。每一个端口都对基本单元，对所用附加译码器以及I/O电路诸方面增加了新的要求。ALU在每个周期中接受两个输入信号而只产生一个输出信号。基于这一点，许多微处理器采用了双端口读出和单端口写入的寄存器堆。

通常是电路规模（指字数）和性能要求规定了单元和读出放大器的类型。对于小的寄存器堆，往往有可能把单个单元作的足够大而且驱动能力足够强，以致于不再需要特殊的读出放大器。当阵列中的单元数增多时，读出放大器的总开销与单元面积成比例下降，而大量单元的成本则上升。使用双端口还是单端口是另一个必须考虑的问题。如果引线间距许可，使用双端口可以节省功耗（写放大器可以作写操作，而读出放大器可以作读操作）并提高速度。然而对有双端或三端口的阵列，布线间距不太允许全部端口都采用双通道。图 8.10 给出了在若干种微处理器中曾用到过的单元结构。

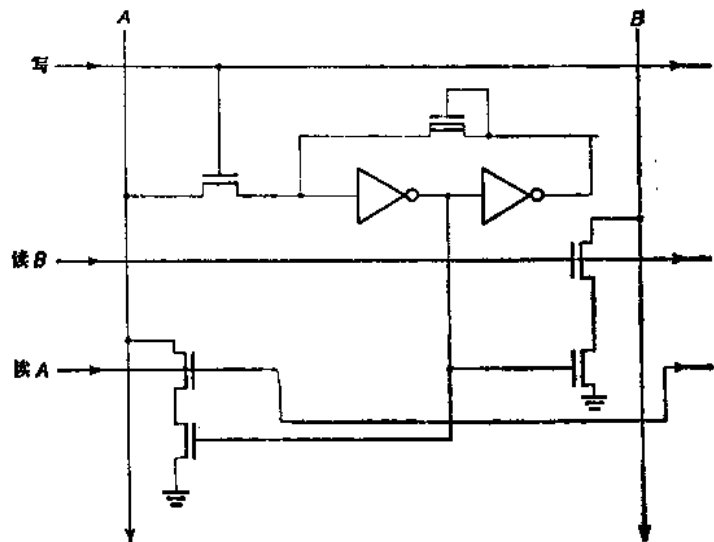


图 8.10 双端口读和单端口写的寄存器单元

片内通信

一个完整的数据通道包括功能部件，寄存器和它们之间的通信总线。总线是数据通道各部件之间实现通信的极有效结构。可惜单总线在一个时刻只能传送一个信号（当然，可以是多个终点）。如果多个功能单元连接到一个共同的总线，它们就不能同时操作，就其性能而言，总线就成了一个薄弱环节。多总线可以改善许多性能。

一种用在所有 MOS 芯片中使互连最少的技术是把逻辑功能块（寄存器和组合逻辑）尽可能做成跟着数据线。使控制线和数据线相互垂直可以得到最有效的版图尺寸。此外，为了避免浪费互连面积，一个区到另一个区数据线的间距必须一致，所以往往是要要求间距最大的部件决定了整个数据构件的间距。对大多数数据通道而言，最少有一个要求每一位有两个输入端的功能单元（例如 ALU）。由此看来，对于 VLSI 数据通道来说，自然应当选用双总线。T11，LSI-11/23，MC68000，HP9000 以及许多其它微处理器都是双总线结构，iAPX432 在指令执行芯片上有一个三总线结构而 NSC16032 则只有一条数据总线。

为了避免浪费面积，如果可能的话，版图最好应有线性结构。在版图中必须同时保证的总线数目是决定占用面积的主要因素。例如，在图 8.11 中，左边和中间的版图分别为双总线和三总线，它们的面积是相等的。右边的版图也具有三总线，但它的面积要大一些。

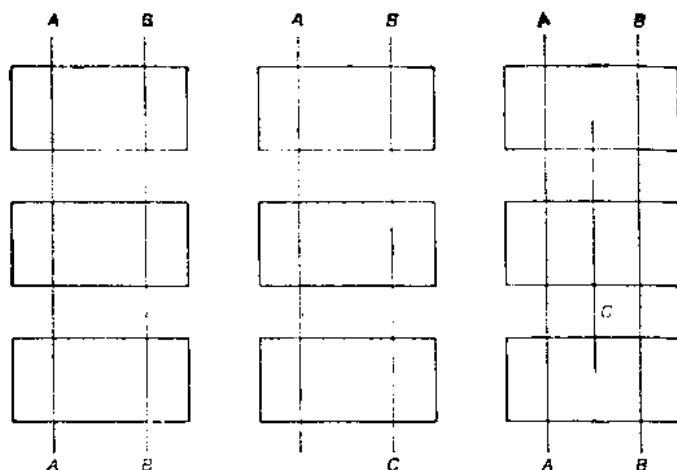


图8.11 三种数据通道的平面布置图

一般讲，对于固定的面积，较多的总线使同时操作的可能性加多。然而由于必须缩短总线就使得通信通道减少了一些（或者是更间接了一些）。在图8.11所示的中间版图中，我们可以同时实现三个局部区域的数据传送。在左边版图中，可以实现两个整体范围内的数据传送，这在中间的版图中就不可能作到这一点。较多的总线还要求来自控制通道的较多信号。

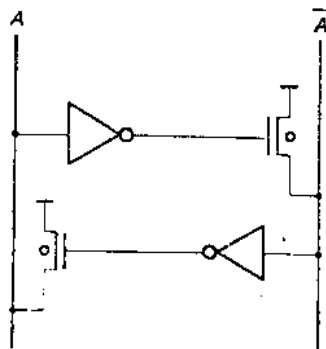


图8.12 在双总线上恢复高电平的电路

如果存在数据间的相关性，在并行操作不可能的情况下，只把功能块加倍是没有用处的。在大多数机器中的寻址操作和数据操作（在8位和16位机器中，它们通常共用同一数据通道）是一对正交函数。对于这些操作，往往使用分开的算术单元。

在微处理器数据通道中，单总线和双总线都用，且总线都是利用预充技术工作的。对于容易受到噪声耦合或电荷衰减影响的总线，或者要求总线能延长一段时间保持其逻辑状态时，可以用图8.12所示技术来使预充总线增强“1”电平。图8.13为MC68000的总线结构。

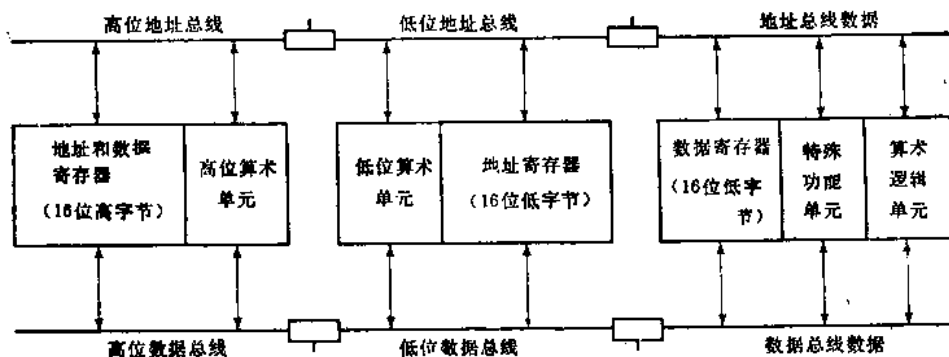


图8.13 MC68000中的数据通道结构。请注意，在各总线之间，是采用传输门连接起来的【Stritter 79】

这种结构在地址单元和数据单元之间可以并发操作，同时还提供了双端口能力而不牺牲通信通道。

例8.1 累加器电路。

我们来分析一个实现连续数据流累加的快速加法器的设计。要作到的关键之点仍是设法避免在反馈环路中的进位传播。在图8.14中给出了一种合理的解决办法。在图内的反馈环路中采用了保留进位加法器，并在环路之外用流水线方式引入进位传播加法器。最高时钟频率决定于通过一个加法器单元的延迟（实际上，最高时钟频率很可能取决于压焊块的 I/O 带宽）。

这种技术可以用于递归数字滤波器。只是在该数据恰被使用前才需要在输出端进行进位传播相加（或相乘）。

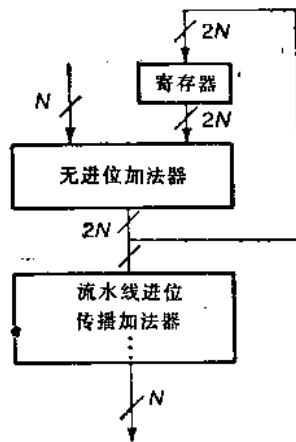


图8.14 长数据流累加所用的电路

8.2 控制结构的实现

控制结构就是通过一系列逻辑状态使机器程序化的逻辑电路。这些状态可具有层次结构，这就是说，主状态 X 可有一些小的相关的子状态 $x_1, x_2, \dots, x_n$ 一直到任何一层。我们把为产生激励信号所需的译码电路也包括在控制部分中。在大多数微处理器中，控制电路和数据电路大体上是等量的。而在另一些数字应用中，整机基本上是一个控制器，往下一直反映到微结构上，这种微结构就几乎整个都是控制电路。反之，专用信号处理器就可能绝大部分是数据通道。

随机逻辑和结构化逻辑是控制部分采用的两种可供选择的主要电路技术，随机逻辑电路是由简单门和存贮元件组成的复杂网络。只有在控制功能比较简单或者是在优化方面尽了很大努力情况下，这种技术才可产生晶体管利用率极高的电路。这种技术的主要

缺点有：增加了逻辑、电路和版图设计的时间，增加了改错工作量，对制造工艺的变化缺乏灵活性。

包括大多数8位机在内的大部分早期微处理器都采用了随机逻辑控制，而大多数16位和32位微处理器则采用了结构化设计。本章的后续部分将专门介绍结构化技术。

规范结构化控制器

我们要讨论的所有结构化控制器，都是由如图8.15所示的组合逻辑和状态寄存器组成的各种规范状态机，包括组合逻辑的各种具体实际电路和状态寄存器的各种组织形式。

具有PLA形式的两级逻辑可以实现任何一般的控制功能。对于结构化逻辑设计和整体效率来说，重要的

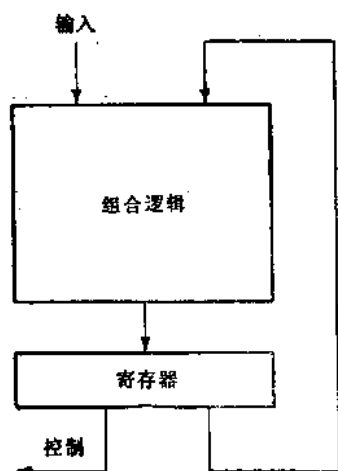


图8.15 有限状态机的组合逻辑和寄存器

是把操作限制为相对少一些的原函数组合，并以这些原函数的操作构成各种复杂的控制功能。这样不仅简化了控制器的设计，而且简化了被控制的数据结构。采用微程序设计可以使版图和数字设计两方面都规整化。对构成复杂控制结构来说，程序设计是已知最强有力的技术。

我们把状态寄存器看作是微指令寄存器（MIR）。状态寄存器的主要变型包括有：不同的寻址时序方法，转移设计方案，状态的层次以及流水线技术。组合逻辑的变型则包括：是 ROM 还是 PLA，是垂直还是水平控制字，以及 ROM 或 PLA 出现的级数和场合。我们将看到，在考虑面积和速度两方面时，这些问题的大部分都可以简化为如何更有效地处理开关序列的问题。

地址排序技术

提供排序功能的简单而又有效的方法是，在微指令字中插入一个次地址（NA）字段。这个字段作为下一个被取出字的指针。如图 8.16（a）所示，通过外界输入端修改次地址字段，就可以实现条件转移。当阵列用 PLA 实现时，就可以在“与”阵列中并入外界输入端，如图 8.16（b）所示。

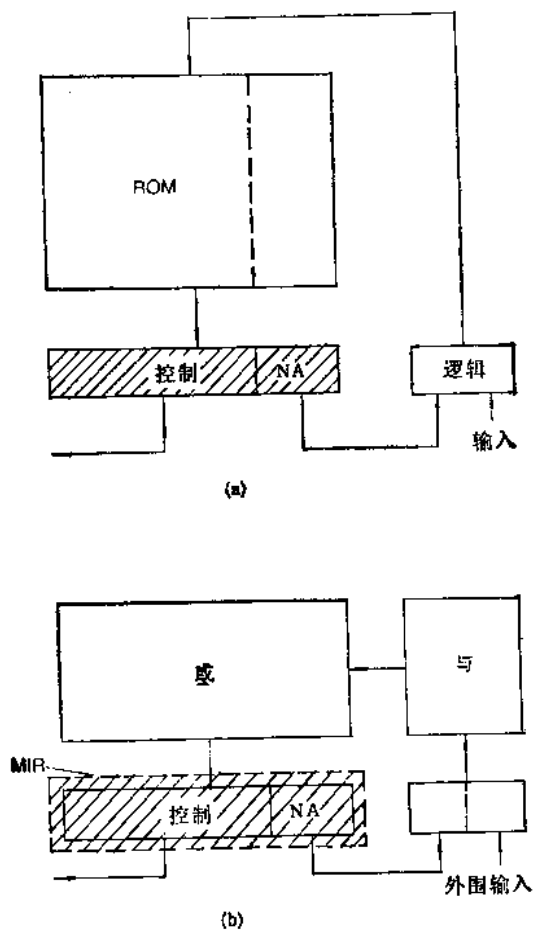


图8.16 次地址字段的采用。(a) 外部逻辑加一个ROM；(b) 加一个PLA。
(a)，(b) 两者都用来构成控制机

在许多应用中，有些情况下在独立的控制通道中往往可能出现微指令的子序列，并且希望只保持子序列的一个拷贝。为了能够使子信息流反过来并入到主信息流中去，就需要提供某种子程序返回机构。由于每个周期信息都要更新，所以把这些信息保存在主控制寄存器中，就不能取得高效率，从而需要某些附加的特殊存储器。虽然可以把这种存储器想像为硬件堆栈，但在许多应用中，简单得多的技术就已经够用了。

在 T11 芯片中采用了次地址字段，宏指令寄存器 (IR) 为微码的多重使用提供了微定序器所必须的转移和返回信息，这样就不再需要其它的连接电路。该结构示于图 8.17。它对于用操作码寻址的操作特别有意义，而这些操作对所有指令都是同样的，而且必须领先于执行实际指令。例如，指令 ADD, SUB 和 EQUAL? 都要求两个操作数，必须从存储器取出这些操作数。在 T11 中，指令型 PLA 决定了要遵循的寻址流，在本情况下，是取出两个操作数的程序。在寻址流的末尾有了两个操作数，机器一定转移到特殊指令的执行流。寻址流的最后一条微指令含有次地址字段，该字段设计成与 IR 一起工作来决定与该条指令¹⁾相应的次地址。

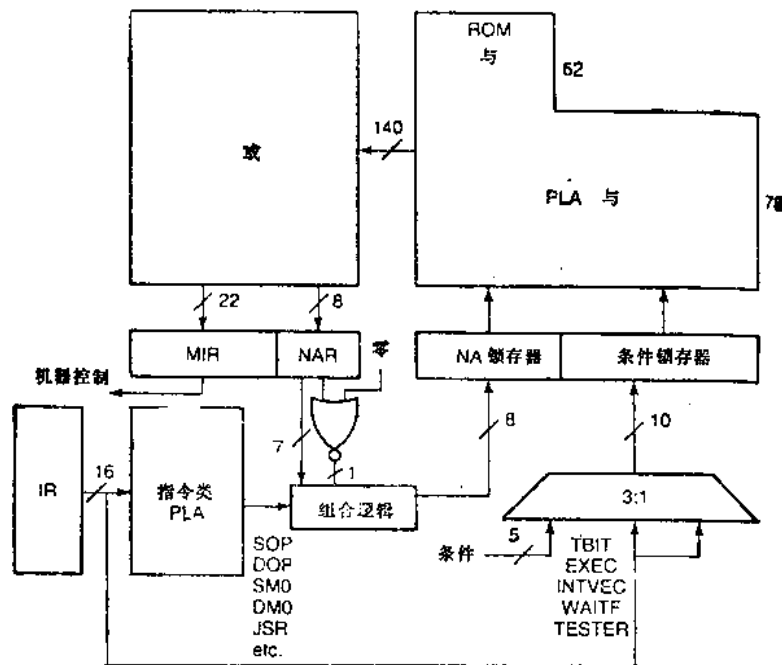


图 8.17 微定序器逻辑 T11

次地址字段的另一个设计方案是采用程序计数器，其主要优点是省去了“或”阵列中的次地址位，节省的位数为 $K \log_2 K$ ，其中 K 是控制存储器中的字数。必须把此优点与采用较复杂定序器和某些时序定时的不利方面进行平衡。一般讲，采用这种方法时，用 ROM 要比用 PLA 更好些。为了提供外界输入端，在定序器中常常需要组合逻辑功能块。如 Mead 和 Conway 所描述 [Mead 80, chapt 6]，在 Cal Tech OM2 中采用了这个方法。在 HP9000CPU 芯片中，显然，使用了一种程序计数器技术，来实现执行 9K, 38 位微指令。如在 OM2 中那样将一组硬件用于子程序返回。

1) 不要把这种技术和结构化的程序设计混为一谈。

有一种修正方案可以改进任一个结构的效率。这就是在外界输入处采用多路开关(缩写为“mux”，多路开关是选择器的另一个写法)。借助于从上一个微码控制多路开关，可以预选决定下一个状态的条件。这就使PLA输入仅限于那些可能影响下一条微指令的信号，从而大大地减少了“与”阵列所要求的输入信号数目。HP9000CPU芯片在整个机器所产生的55个独立条件之间进行选择。多路开关作为总线的一部分遍布于整个机器中。在逻辑领域中，我们扩展逻辑级数超过要求的两级，以便使面积最小。这类控制总线由数据线和控制线组成，它们决定了哪一些信号源对于某一个周期应是有效的。控制线可能沿着数据总线布线，或者有可能内含在微指令码中。

条件转移设计

我们已经接触到了功能转移或非条件转移。因为定时电路可能在机器的关键通路中，所以条件转移是一种特例。大多数微程序化的机器至少取微指令与执行数据通道是重叠的，如图8.18所示。这样一来，当数据通道正在执行由第 N 条微指令控制的操作时，控制通道则正在取出数据通道操作所需的第 $N+1$ 条微指令。

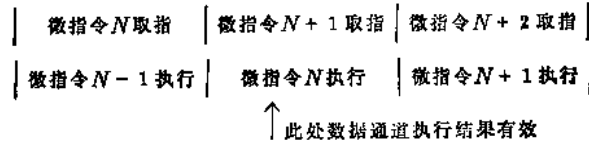


图8.18 重叠的微指令取指和执行定时

从图可以看出，取决于数据通道第 N 条微指令操作结果的条件转移并不会影响第 $N+1$ 条微指令的取指，这是因为这二者是同时进行的。这样，直到第 $N+2$ 条微指令被执行之前，条件转移不大会生效。这就导致程序中存储区 $N+1$ 也许不得不白白浪费的。对于一台普通的计算机，这个论点无疑是正确的。然而，对于LSI，我们总是可以利用列选择比在ROM中选择一个新行的操作¹⁾快得多这一特点，得到一个简单的窍门。若把微存储ROM组织成每次存取使列选择器有两个或更多个待选指令码，当条件转移限定在列选择器级进行时，实际进行选择的时刻就被延迟到周期的后期。曾报道过莫托罗拉68000和HP9000微计算机采用了这种设计。MC68000有四个字送到多路开关作为转移选择对象。而HP机器就有32个字送到多路开关。在图8.19中以一个四选

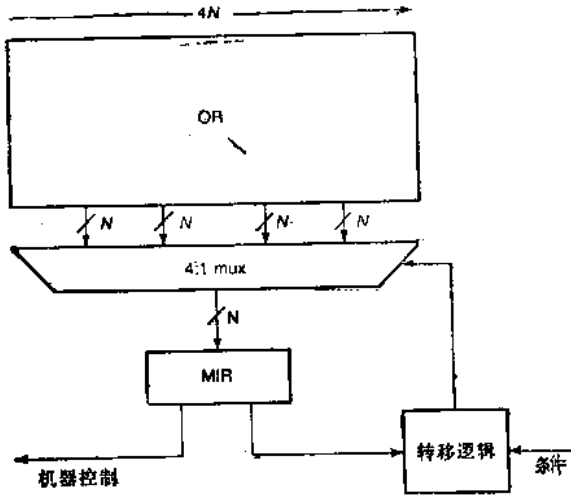


图8.19 “页”式ROM寻址。其中转移逻辑从四条可能的微指令中选择出加到MIR中去的指令

路开关作为转移选择对象。而HP机器就有32个字送到多路开关。在图8.19中以一个四选

多路开关作为转移选择对象。而HP机器就有32个字送到多路开关。在图8.19中以一个四选

1) 常常把它叫做“页模式”寻址。“行”定义了不同的页，而“列”则规定了它在页中的位置。

在不同的控制层次上,使条件转移影响最小的另一个办法是假定转移沿着一条最可能的通路。例如,考虑在一个程序循环中的转移,假定转移的终点就是程序循环的起点,对多数分支来说不会带来什么损失。

水平控制码与垂直控制码的比较

水平方式和垂直方式之间的区别与微指令的位宽度有关。最低一级的控制功能仍是二进制控制点,它是一条送到一个基本逻辑功能块的 GO/NOGO 命令。归根结底,所有控制码都必须完成这些基本任务。在全水平控制码中,每个基本控制点为一位。近代的微处理器大约有 100 到 300 个控制点,这就意味着 MIR 的宽度要有 100 至 300 位。如果位的所有组合都用上,则不可能再进行别的编码了。幸好事实并非如此,所以仍编码可用。(对于一个 10MHz 的微处理器,用一个宽度为 100 位的 MIR 来运行所有可能的组合将需要 4×10^{15} 年)。在水平/垂直区别上,常常正是这种编码程度问题。绝大多数计算机和实际上所有的微处理器都把微码再细分为对应于控制各个逻辑单元的控制字段。例如,细分为寄存器堆、ALU 和总线等等控制字段。每一个单元都有它自己的译码器,用以把控制字段转换为各基本控制信号。在一个微指令系统中,可同时存在的字段数常常是重要的设计参数。如果每一个字段控制一个逻辑单元,则字段数决定于能同时操作的最大单元数。这一特征再加上与计算机内部供数据操作的总线,决定了在微码级上可得到的最多并行操作。

微定序器的最终垂直编码是使所有字段压缩成的一个二进制编码字段。毫微码(nanocode)ROM 可以用这微码为计算机在 V 种可能控制码中选出一种,如图 8.20 所示。这样一来,微码 ROM 就有 $\log_2 V$ 个输出,它可以在毫微码 ROM 中寻址 V 个字。这些字要求最快的再译码。在计算机中,微码/毫微码结构是非常普遍的。早在 LSI-11 时就把它们用在微处理器中了。在毫微码 ROM 中可以使用反馈,从而使它对微码 ROM 输

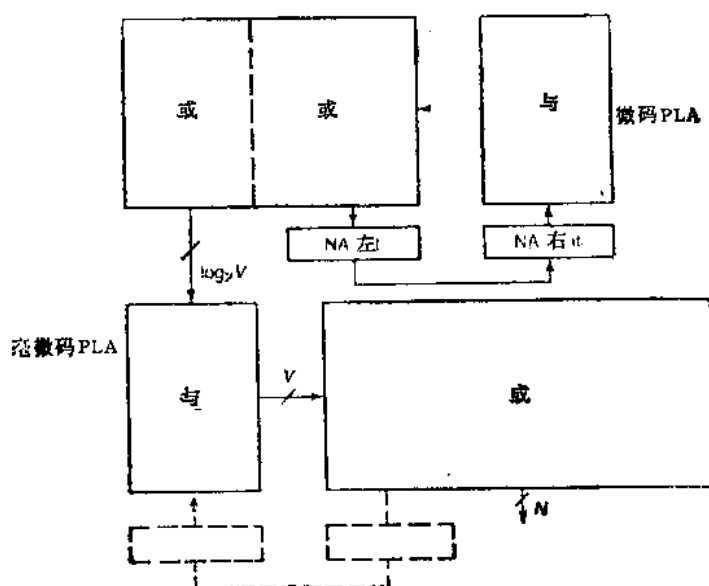


图 8.20 二级微码/毫微码控制通道

出作解释而不仅仅是对它译码。在希望使用很宽的控制字时, 这种技术就有不少好处。

在毫微码ROM中不用反馈的时候, 可以把一级电路和二级电路的大小作一比较。图8.21给出了一个具有 K 个 N 位的字的简单微码PLA。一级电路设计所占用的面积 A_1 大约为:

$$A_1 = K(N + 3\log_2 K) \quad (8.5)$$

如图8.22所示, 在二级电路中, 电路所占面积 A_2 为:

$$A_2 = K(\log_2 V + 3\log_2 K) + V(2\log_2 V + N) \quad (8.6)$$

共享比 ρ 定义为: 毫微存贮器中的字被微存贮器中的字所共享的平均倍数,

$$\rho = \frac{K}{V} \quad (8.7)$$

在微码PLA中 $K-V$ 输出的字是重复的。重写 A_2 , 我们有

$$A_2 = K(4\log_2 K - \log_2 \rho) + \frac{K}{\rho} \left(N + 2\log_2 \frac{K}{\rho} \right) \quad (8.8)$$

在 $A_1 = A_2$ 时, N 的值由下式给出:

$$N = \frac{\rho + 2}{\rho + 1} \log_2 \frac{K}{\rho} \quad (8.9)$$

当 ρ 较低时, 在这种技术中, 要使面积利用更为有效 N 一定是大的。例如, 对 ρ 值为1.4和2, N 值分别为70和32, 它们都有 $K=512$ 字。微码/毫微码结构的功耗来自具有极宽控制字的存贮器, 尽管这些字在多重控制流中使用, 但在毫微存贮器中只用一次。

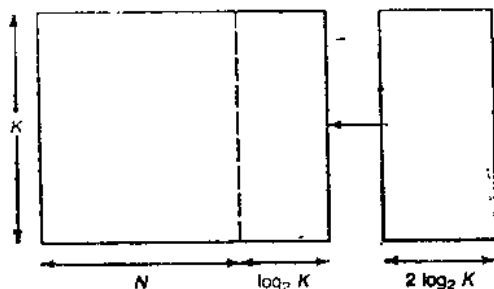


图8.21 一级微程序控制结构

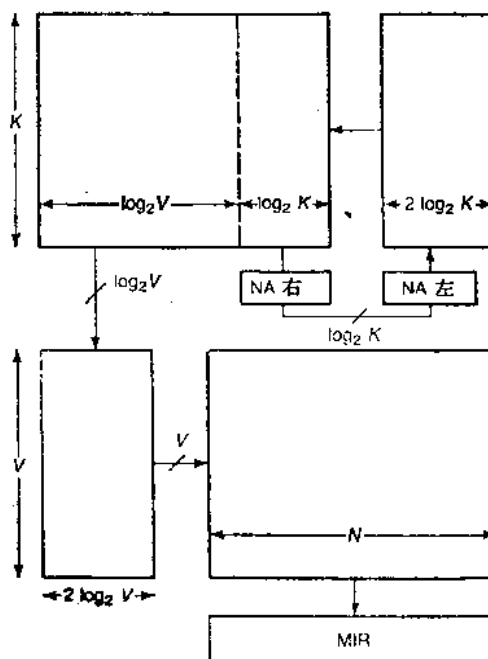


图8.22 与图8.21具有同样指令数和同样MIR宽度的二级毫微码结构

MC68000中采用了这种技术的一种变型。由于其中的阵列是组合在一起的, 所以MC68000中的结构是比较复杂的, 这种组合使加到第一个PLA与阵列中的输入信号可以在毫微存贮器中直接寻址。图8.23说明了类似于MC68000的结构。请注意, 在这种结构中是怎样用选择器去匹配毫微阵列和微阵列的宽度的。

对毫微存贮器要考虑的最后一点是集中译码和分散译码之间的区别。我们已经看

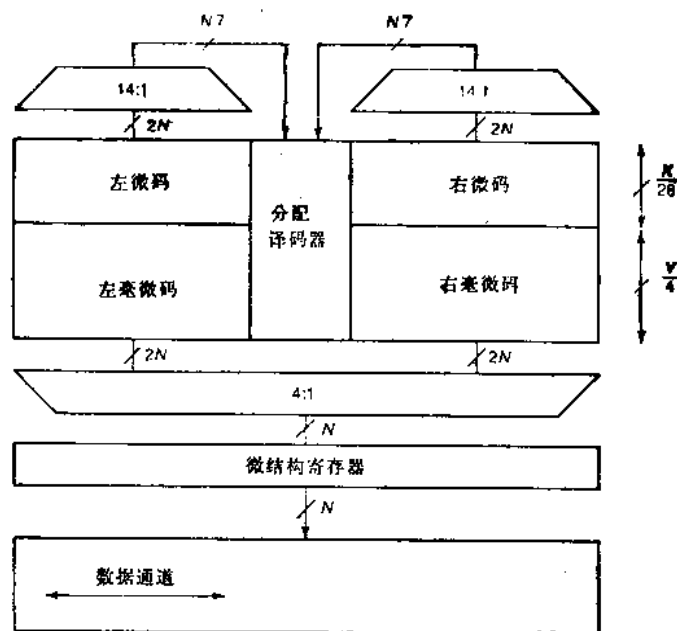


图8.28 MC68000 中的控制结构

到，只有在输出具有相对地宽(极为水平的)的控制字时使用毫微存贮器才是合理的。这就意味着在大量布线从毫微存贮器引出，而在布线终点只作少量附加译码。因而对译码器来说毫微存贮器具有阵列效率高的优点，而其缺点是有更多的全局布线通过整个芯片。何种方案能得到真正的好处将由特定的条件来确定。

状态的分层结构

除微指令寄存器以外，往往要维护一些控制状态。在微处理器中，这起码有宏指令寄存器。以前我们曾提到，用IR来提供一种上下文关系，使共享的微码子流程能够调度到正确的下一个地址。IR提供了控制辅助状态排序的主状态信息。主状态信息还可以控制由辅助状态排序所产生的触发信号的译码。再者，用IR作为主状态寄存器，可以看到主状态是如何直接影响机器控制的。在指令寻址格式是固定的计算机中，像PDP-11或MC68000，其宏指令的地址字段，可以用来直接对机器中的局部寄存器寻址。如图

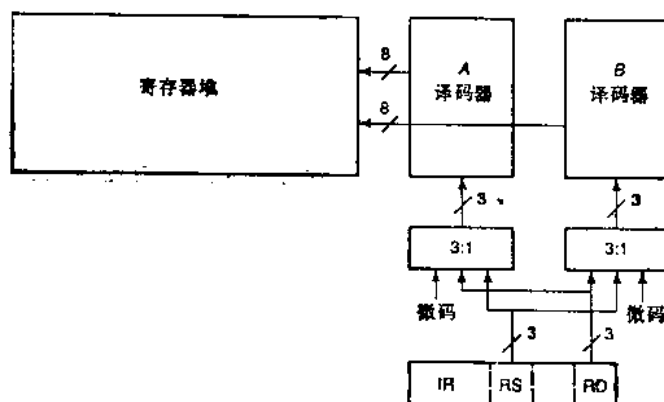


图8.24 具有局部译码的寄存器堆节省了ROM空间

8.24 所示, 微指令仅提供了一个从宏指令特定字段到寄存器堆译码器之间的连接通道。在第5.9节中曾讨论了这种技术, 这里的问题是简化信息通信和缩短临界通路。

在毫微码操作对微码指令进行解释的情况下, 本节中前面提过的微码/毫微码结构是另一种分层结构状态的示例。

对所有宏指令通用的主操作集有

1. 取出指令,
2. 操作译码,
3. 取操作数 (可能是多重的),
4. 执行指令。

这些主操作可等同于机器的主状态。因而除宏指令寄存器以外, 设置一个跟踪当前主状态的寄存器往往是有好处的。为了多重操作指令的微码可以共用取操作微码, 就要使用一个计数器以区分第一操作码和第二操作码等等。对于执行诸如乘和除这种要求微指令循环的操作来说, 环路计数器是很有用的。问题的自然分层和分割往往会启发出很好的硬件装置, 这是本书的经验。

控制通道中的流水线作业

一般讲, MOS 逻辑电路适于采用流水线技术。这是由于在 MOS 电路中, 加入数据锁存器只增加很少的成本。在增加工作容量方面流水线是极端有效的, 它可以在不增加并行硬件的情况下实现同时操作。其主要缺点是在不同级流水线之间数据相互依赖时, 就会带来定时问题。如果使用独立的硬件模块来替代流水线, 则这些定时问题就表现为连线问题, 并且在流水线需要充分利用的情况下, 大多数硬件模块就会闲置。我们看过一个微指令取指和指令执行有交叠的流水线简例, 其结果导致了条件转移。

绝大多数高速计算机采用了预取指令流水线, 近代的高性能微处理器都是这样设计的。为了减小控制存储器的容量并允许交叠操作, 使用了独立的变换 PLA, 以寻找每条宏指令的起始微码地址。因此, 预取指流水线大体上如下:

1. 预取宏指令,
2. 用变换 PLA 找出微码起始地址,
3. 取出微指令,
4. 执行指令。

并不是所有操作都要占用相同的时间 (特别是预取指本身), 但多于四个微周期时间大概是必要的。一条宏指令的转移要求整个流水线激励和重新启动。在一些机器中, 平均讲每 3 至 4 条指令就要出现一条宏指令转移。这是因为在硬件中已经设置了较多的并行通道, 使得转移变得更加普遍了。对于转移, 统计的预取指算法在保持预取指的高性能方面是特别有用的。

把预取指单元看做是半自主的机构往往是有用的, 它尽可能经常地访问存储器, 对 (也许) 未用的指令取指, 然后把这些未执行的指令排成队供执行单元使用。事先不能预料的转移要求对该排队进行激励。

为了从预取指令中得到最佳性能, 就要求存储器具有很高的带宽。一种可能性是使存储器操作本身是流水线的, 使得取指和存储器中的数据操作可以同时进行。在 HP

9000 系列的 CPU 和专用存储器芯片中,集成了流水线存储器结构。紧密耦合的高速缓冲存储器是另一种可能性。这样就避免了流水线方法的某些复杂性。在主机中,这两种技术都很有用。对于集成度非常高的芯片,片内高速缓冲存储器极具吸引力,这是因为它可以从大多数操作的关键通路中去掉潜在的片外延时。Z80000 报道了片内程序和数据缓冲存储器的使用。MC68020 报道了片内程序缓冲存储器的使用。

用于微处理器组合逻辑的ROM和PLA的比较

在用户定制 MOS 电路中,对于微处理器来说,可以自由地选用 ROM 或 PLA,每一种都各有利弊,但都已用于微处理器电路。正如我们将来看到的那样,在小规模的控制应用中,一般讲 PLA 具有优势,而 ROM 更适于大规模电路。首先我们考虑一下二者的实际电路和版图。

从逻辑上讲,可以把 ROM 看作是“与”阵列全编程的 PLA 的一个子集(不是无条件的)。因为 ROM 的与阵列是全译码,所以和 PLA 的与阵列相比,ROM 的与阵列利用晶体管更为有效。在总效率方面这是一个要考虑的因素。然而 ROM 往往要求有一个外电路,以便对输入进行编码,从而与二进制兼容。这一附加的编码会耗费时间和面积。对于小的阵列,总的来说往往 PLA 中的晶体管较为有效。在大阵列中,ROM 与阵列所节省下来的面积可以超过输入编码电路的面积。然而,延时损失仍旧存在。

在 LSI-11/23 和 T11 中曾经引用了一种技术,用它可以减小 PLA 的无效性而同时保持了两级逻辑结构。这些芯片采用了次地址定序设计,它与修改量(称为翻译器)一起提供了转移操作。只有作为转移目的地的指令才被存在 PLA 中,其余部分则放在用 NA 字段作为二进制编码地址的纯 ROM 中。最终的物理结构可以互相渗透在一起,在 T11 芯片中就是这样作的。图 8.17 中说明了这种技术。

在结构上,ROM 确实为设计者提供了极好的选择余地。ROM 译码可以在行与列间任意进行划分,在芯片版图设计中,这一点可能是重要的。同时,正如本节前面讨论过,转移的选择可以推迟到列译码时刻,在这种情况下,这一点对于使条件转移的影响减至最小是有益的。

例 8.2 连接计算机(The Connection Machine)。

曾经作过许多努力,企图通过采用新结构来克服冯·诺曼存储器带宽这个薄弱环节。一种常用的方案是把微处理器分散到存储器的各个部分,使存储器变成所谓“智能存储器”。MIT 连接计算机就是其中一例(Hillis 81)。它含有大量的小微处理器。每一个微处理器都有略小于 1K 位的存储器和一位串行 ALU。从我们的观点来看,一个明显的问题是由于 Carl Feynman 引起的互连设计问题。当连接计算机有一个总体通信机构时,它被想像成计算机形成一个巨大的网格来解某些构造良好的题目。数据流沿着四个方向中的任一个方向传送。然而,由于每个芯片含有 64 个微处理器,所以每个芯片的引出管脚数将会非常之多。曾经开发了一个变通的设计,它少用了 25% 的引出管脚。在图 8.25 中给出了一个常规的连线方案。在同一时刻如果限制数据只在一个方向上流动,这样在任一时刻就只使用一半连线。图 8.26 说明了连接计算机的技术。所有驱动器都是三态 I/O 型。图 8.27 表明了数据是怎样由南向北流动的。这是拓扑关系怎样在很高的层次上影响计算机结构的一个示例。

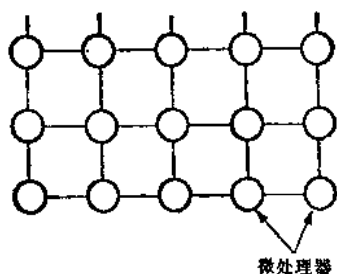


图8.25 微处理器矩阵的
常规连线图

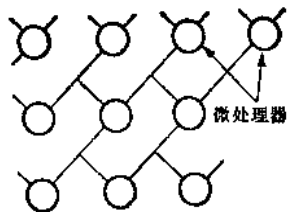


图8.26 连接计算机的微处理
器组织方式

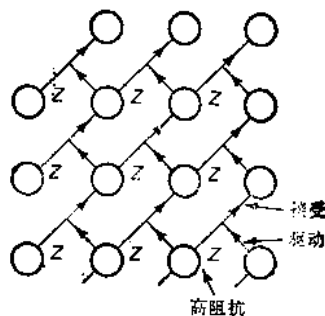


图8.27 连接计算机的数据由南向
北移动的示意图

8.3 测试和故障排除

在 VLSI 电路故障被检出、验证和成功地转向生产之前，设计还是没有完成的。由于设计时广泛使用 CAD 工具，使得这项任务不太困难了。基本工具包括电路模拟器、开关或逻辑级模拟器、寄存器级模拟器、设计规则检验器和连续性检验器。曾经编写了许多附加的程序以提高设计者的生产率和成功的几率。这些程序对以下各程序块进行寻址：工艺程序、器件建模、符号版图压缩、自动布局和布线、设计图输入、电路拓扑关系和参量提取、逻辑验证、方案验证、定时分析、电路优化、逻辑优化、平面布局、微码汇编和编译、统计模拟、收缩阵列优化等等。实际上，一个更有挑战性的设计自动化目标是把这些程序和低层次的数据库统一到一起，从而把它们合并为一个问题。尽管芯片复杂度一再增加，但越来越多的芯片是一次设计成功的，这证实了上面那些在 CAD 方面研究的成果。

即使芯片设计正确，由于一定比例的芯片有加工缺陷而导致成品率下降，必须从坏的芯片中把好的挑选出来。在测试阶段必然提出两个截然不同的问题。第一个问题是故障检验，这是指分析芯片的设计错误。第二个问题，在制造界习惯上把它叫作测量，通过它可发现制造工艺中的错误。虽然二者都是从预见中受益，但这两个问题却是十分不同的。如果已设计的芯片没有考虑到故障检验和测试任务，那么在这两个方面都将很可能会落空。在设计芯片时就把测量考虑在内是可能的，这就是通常所说的可测性设计。

故障排除

当第一批芯片从生产线上制造出来时，并不知道它们存在的问题，但这些问题可以是设计错误，制造偏离和加工缺陷。第一步是减少某些严重制造错误的可能性，这可以通过检查预先放在芯片上的参量测试图案来实现。工艺测试图案(NBS 83)起码应包括：

- 各种类型和不同尺寸的晶体管。
- 扩散结构的击穿和电容。
- 所有氧化物和电极材料的电容器，以检查电介质的厚度和完整性。
- 接触孔长链，以确定接触的完整性和接触电阻。

- 所有信号导体的电阻图形，包括金属等等。
- 环振或其它性能测试图形，设计它们是为了将工艺特性综合为性能优良。

如果已经查明工艺差不多是全对的，下一步是找出没有制造缺陷的芯片。最好是从硅片不同部分（以减少某种类型的掩模缺陷）找出两个功能相同的芯片。如果这些芯片功能正确，则故障排除就完成了。否则就必须进入故障排除周期。故障排除的主要目的是找出每层掩模的最大可能错误数目。错误往往是相互套在一起的。例如，电源 V_{DD} 和地线的某一处短路，会掩盖两个信号线互相交叉的错误。消除 V_{DD} 和地的短路是极端重要的，这样才能发现第二类错误。而信号线交叉又会掩盖微码故障。每次返工制作新的掩模，都要损失大量时间。下列技术在故障排除周期中是很有用的：

- 利用超声切割器或激光粉碎器切断连线。这就能把芯片的某一部分电路隔开并去掉短路。
- 用探针探测芯片内部点。探针的电容要小到 0.1pF 至 0.01pF 才行。有可能探测芯片上任意一条可见的金属线。也可能探测多晶硅线。但是这是十分困难的。重要的是把所有准备检测的节点都引到上层金属层（另一种方法是用电子束机以电压衬度模式扫描芯片）。
- 注入信号，这样作可以把电路中没有功能的部分旁路掉从而找出相互套在一起的故障。
- 在每一个主要工艺步骤之前保留片子，这就可以很快地找出下一层掩模的错误（工艺错误或设计错误）。
- 在关键节点处安放漏级断开的晶体管，如图 8.28 所示。其漏端要引到最上边的金属层，它的源应当接地。在该节点外加一个低阻抗上拉元件，即使是用探头输入电容为 2pF 的示波器（例如 $5\text{mV}/\text{度}$ ）也可以观察到非常高速的电压波形。例如， 2pF 电容和 100Ω 负载电阻的 RC 时间常数为 0.2ns 。
- 利用静态或半静态锁存器使电路单节拍动作。
- 采用特殊的故障检测电路。例如，可以直接进入到控制 ROM 强迫寻址。
- 利用工艺偏离，故意出现工艺参数最坏条件，以便很快地诊断出由于勉强够格的设计而带来的问题。

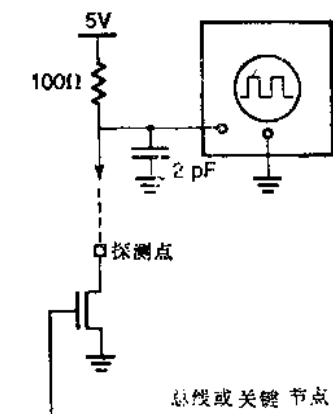


图 8.28 探测高速总线技术

生产测量

生产测量的目的是找出最简单的输入激励信号组合，用以表明每一种可能故障都已成功地避免了。这里所谓的故障是指逻辑或开关一级的。例如，制造工艺中的缺陷可以使某些物理元件失效，从而表现为逻辑故障。故障模型是指理想化的故障，用它从形式上探讨一些有关问题。最普遍的故障模型为“钉住(stuck-at)”模型。在单一钉住模型中，假定某一电路失去功能，其原因是电路中的某一个节点被固定在“1”或“0”状态。这个模型显然存在很大限制。例如，它没有包含两个信号线短路在一起的情况。然而经

验表明, 如果一个 LSI 系统能够通过一种为找出绝大部分单一 Stuck-at 故障而设计的测试, 则芯片就可能在各个方面都是好的。在越来越高的集成度下, 这是否仍旧真实? 这只能由时间来回答了。

在硅片还是整片时, 对 VLSI 芯片进行测量是最经济的手段。探针加到每个芯片上并加一个激励矢量。为了减少测量时间, 把测量安排为首先找出最容易发现的故障。一般讲, 严重的参数如电流 I_{DD} 和管脚漏电等要首先进行测量。当芯片在某一测量失效时, 就用一滴墨水把它标出来, 以便在随后切片时, 把它们识别出来并除掉。VLSI 芯片的整套测量所耗费的时间, 反映到最终产品成本中可能占有很大的比例。测量 LSI 芯片每秒钟要耗资几美分。尽管如此, 进行测量往往还是比不测量便宜得多。据分析, 在封装好的集成电路中检测出一个故障的耗资是在芯片上查出一个故障的十倍。在印制电路板上查出该故障的耗资则需再加十倍。而在系统一级上查出故障的耗资还得再加十倍。假如一个有故障的系统被送出去了, 找出并矫正该故障的耗资又要高一个数量级。显然经济性要求早点检出故障。测量所需的时间大体上是按芯片所含晶体管数呈线性律和平方律之间的某个规律变化的。

测量的第一步是选择一组随机测试输入信号或一组功能测试信号。在组合逻辑芯片中, 随机产生测量图案可以检出 55—90% 的故障; 然而, 对于不内含扫描技术的微处理器类电路, 采用随机产生图案只能查出 20—30% 的故障。对于这类电路, 常采用反映该电路功能规范的测量图案。这些测量图案可以查出大约 75—90% 的错误。测量程序扩大到包括因屏蔽不良 [Goel 84] 带来的字段错误, 将使上述百分比随之增加。

可测性设计的基本点是可观察性和可控制性。对于电路中的每一个节点, 都有 Stuck-at 失效的可能。为了测出一个节点钉在 1 和钉在 0, 就必须控制该节点使它置为高和低电平。节点电位难以控制, 测量它就更难了。例如, 为了把计数器的高位置位, 就要求输入一个非常长的序列, 因此很难控制这些高位码。为了控制某个内节点, 要确定一条从输入到该节点的灵敏道路。这就要求将该通路上各种逻辑门的输入端设置到相应的逻辑值。(送到与非门的“其它”输入端接到高电平, 送到或非门的“其它”输入端接到低电平)。另一半问题就是可观察性。我们不仅要控制每个节点, 而且要观察我们设定的节点值。还必须找出连到输出端的灵敏通路(它与输入所要求的灵敏通路并不矛盾)。寻找组合逻辑块正确输入信号的算法包括有 D 算法 [Roth 80] 和 PODEM [Goel 82]。太难测的节点就留下来不测了。

建立一个好的芯片测量程序, 包括产生测量图案和验证测试两个方面。验证是通过重复故障模拟而进行的。故障模拟包括在电路模型中插入一个故障, 然后论证该测量矢量确能找出这个故障。必须针对每一种可能的故障都这样作。完成这些任务所耗费的计算机时, 大约按 N^K 变化, 这里 $2 \leq K \leq 3$, N 为晶体管数。如果是时序电路, 能处理的最大实际电路大约为 2000 门。如果是组合逻辑电路, 能处理的最大实际电路为 5000 门。因此, 必须把整个电路分成一些独立的小功能块, 而如果这些功能块是纯组合逻辑电路就最好不过了。MC68000 级电路的测量验证用小型计算机需一周的机时。

设计一个电路必须是可测的, 这往往要求设计方法的某些折衷。实现这一点有常规的和特设的两种技术。关键在于把整个电路分为一些小的组合逻辑功能块, 使得其中的输入端和输出端都很容易控制和观察。这包括插入附加测试点或测量模式。比如, 对微处

理器来说,在较高层次上容易控制和观察 MIR 中的内容显然是重要的。所有 ROM 和 PLA 中的内容一定是易于读出的。要能控制和观察所有 VLSI 芯片中寄存器的内容,乃是可测性设计的一个重要部分。特征分析和扫描设计是使一个芯片可测的两个常用技术。

在扫描设计中,把所有 VLSI 中的寄存器(除去那些易于控制和观察的,比如在数据通道中的那些寄存器)串联成一个长的移位寄存器,然后使之工作在测量模式。这就使电路其余部分简化为组合逻辑块了。移入输入信号(作为控制),系统受时钟定时,然后输出值就移出(作为观察)。这类系统的测量图案比较好生成。这种技术要多占用 4—20% 的芯片面积。对于采用动态存储器的电路(一般 MOS 电路都如此)不利于扫描设计。图 8.29 给出一种可能的扫描寄存器。

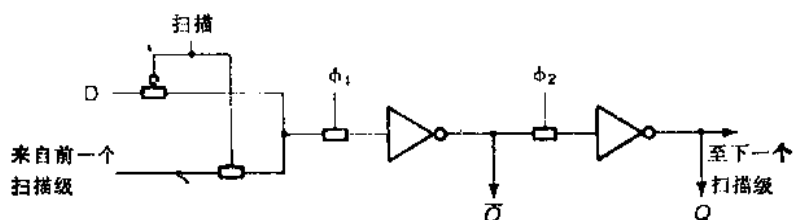


图 8.29 采用动态逻辑的扫描寄存器

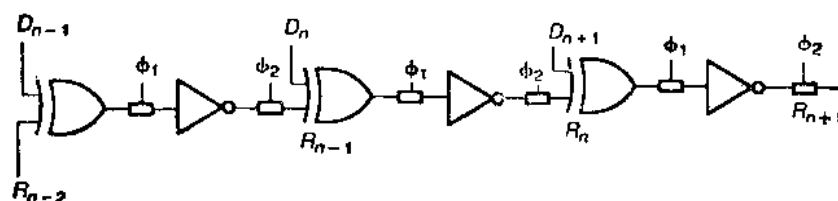


图 8.30 用来采集测试点 D_n 特征的电路

在特征分析中,开发了伪随机序列发生器,如例 1.4 所示。该序列发生器可以实现两种功能。它可以对那些从外部管脚难以寻址的内部节点提供随机输入信息。如图 8.30 所示的改进电路还可以采集从其它发生器或从外部输入引起的一些特征。当采集特征时,输出信息往往是以位串行方式连续地移位出去的。

其它许多技术均是上述技术的变型或组合。有兴趣的读者可以参考文献。

8.4 小 结

一个设计良好的 VLSI 反映了电路装置的能力。绝大多数 VLSI 系统可以分块为数据通道和控制通道。在控制通道中,关键设计决策是围绕着采用阵列结构的种类、控制集中的程度和流水线化的程度诸方面进行的。在数据通道中,关键的决策包括功能单元的数量和类型,以及连接它们的总线数量和类型。在分割一个 VLSI 系统时,某个功能块属于哪一部分往往是不明显的,解决这些问题的关键在于考察各种可选方法是如何实现通信的。高性能意味着控制通道和数据通道间的紧密耦合——这就是收缩系统效果很好的理由之一。

当我们花了大量精力去研究优化时，请记住 MOSVLSI 的实际功耗是比较经济的。在与其它技术作比较时，就纯速度而言，MOS 不算快，但是 MOS 的单位成本的速度比是优越的。MOS VLSI 的功耗决定于设计师解决一个双重问题所具有的技巧和创造性。这两个问题就是：几百万个晶体管几何布局的物理问题以及利用这些晶体管能力的策略问题。

习 题

- 8.1 试说明先行进位加法器中复杂门电路所要求的下拉晶体管串，近似等效于预充曼彻斯特进位链电路中所需的传输晶体管链。
- 8.2 在 CMOS 组合逻辑电路中，一定类型的故障可使一个组合逻辑网络转变为时序网络。试给出一个这种故障的示例。
- 8.3 我们在 8.2 节中讨论过，将 ROM 用在“负”模式中是怎样有助于系统适应于条件转移的。试解释负模式 ROM 怎样寻址可以加速执行直接式程序。
- 8.4 对于一个给定的指令系列，微处理器的性能与晶体管的数量之间的关系表现为类似图 8.31 所示的曲线关系（假定这些晶体管的使用都是恰当的）。请注意微码存储器（假定是采用的）和任何局部寄存器以及高速度缓冲存

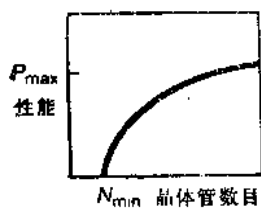


图 8.31

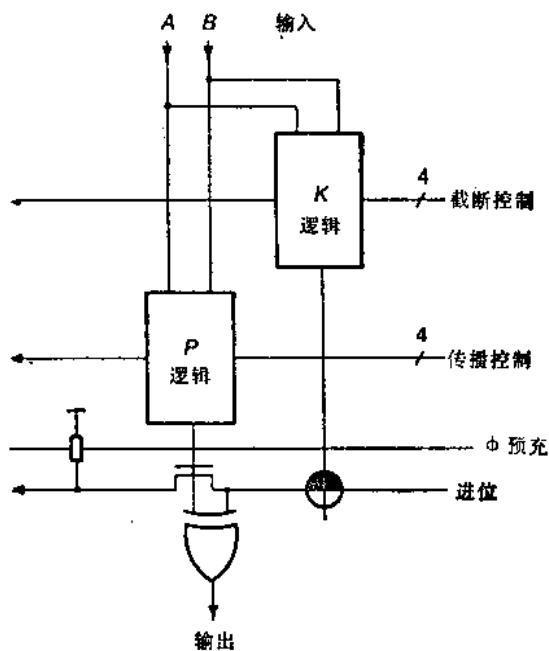


图 8.32

贮器要包含在计算之内。试考查 N_{min} 和 P_{max} 的起因。ΔΔ

- 8.5 在二级微码/毫微码的计算中，我们假定用一个 PLA 作为毫微码存储器。若假定毫微码被存在一个 ROM 阵列中，试重作这些运算。
- 8.6 试重作示例 8.1 的累加器电路设计。设计环路中要具有流水线进位传播加法器，且使它的速度恰好与示例中保留进位电路的速度一样快。
- 8.7 图 8.32 说明了一个简单的一位片式 ALU。它采用了两个功能块。试解释怎样用这个 ALU 实现 $A \cdot B$ 逻辑操作。试解释怎样实现算术二进制补码操作 $A - B$ ？
- 8.8 假如有一个预充曼彻斯特进位链。试说明在 32 位加法器的第 15 与第 16 位之间，有效进位信号的传播几率为 0.25（假定是随机输入）。

参 考 文 献

[Alpert 83] D. Alpert, D. Carberry, M. Yamamura, and P. Mak, "32-bit Processor Chip Integrates Major System Functions," *Electronics* 56, 113-119, July 14, 1983. (The Z80000)

- [Anceau 80] F. Anceau, "Architecture and Design of Von Neumann Microprocessors," NATO Advanced Study Institutes Series, series E., no. 47, 301-327, July, 1980.
- [Bayliss 81] J. A. Bayliss, J. A. Deetz, C-K Ng, S. A. Ogilvie, C. B. Peterson, and D. K. Wilde, "The Interface Processor for the Intel VLSI 432 32-Bit Computer," *IEEE J. Solid-State Circuits* SC-16, 522-530, 1981. (The i43203)
- [Bayliss 81b] J. A. Bayliss, S. R. Colley, R. H. Kravitz, G. A. McCormick, W. S. Richardson, D. K. Wilde, and L. L. Wittmer, "The Instruction Decoding Unit for the VLSI 432 General Data Processor," *IEEE J. Solid-State Circuits* SC-16, 531-537, 1981. (The i43201)
- [Beck 84] J. Beck, D. Dobberpuhl, M. J. Doherty, E. Dorenkamp, B. Grondalski, D. Grondalski, K. Henry, M. Miller, B. Supnik, S. Thierauf, and R. Witek, "A 32b Microprocessor with On-Chip Virtual Memory Management," *IEEE International Solid-State Circuits Conf.*, 178-179, San Francisco, Calif., 1984. (A single chip VAX)
- [Bell 71] C. G. Bell and A. Newell, *Computer Structures, Readings and Examples*, McGraw-Hill, New York, 1971.
- [Bell 78] C. G. Bell, J. C. Mudge, and J. E. McNamara, *Computer Engineering*, Digital Press, New York, 1978.
- [Bennetts 82] R. G. Bennetts, *Introduction to Digital Board Testing*, Crane Russak, New York, 1982.
- [Bennetts 84] R. G. Bennetts, *Design of Testable Logic Circuits*, Addison-Wesley, Reading, Mass., 1984.
- [Beyers 81] J. W. Beyers, L. J. Dohse, J. P. Fucetola, R. L. Kochis, C. G. Lob, G. L. Taylor, and E. R. Zeller, "A 32-Bit VLSI CPU Chip," *IEEE J. Solid-State Circuits* SC-16, 537-542, 1981. (The HP 9000 CPU)
- [Bhavsar 81] D. K. Bhavsar and R. W. Heckelman, "Self-Testing by Polynomial Division," *IEEE Test Conference*, 208-216, Philadelphia, Pa., 1981.
- [Budde 81] D. L. Budde, S. R. Colley, S. I. Domenik, A. L., Goodman, J. D. Howard, and M. T. Imel, "The Execution Unit for the VLSI 432 General Data Processor," *IEEE J. Solid-State Circuits* SC-16, 514-521, 1981. (The i43202)
- [Byrne 80] W. J. Byrne, "Three Decapsulation Methods for Epoxy Novalac Type Packages," *Proc. IEEE International Reliability Physics Symposium*, 107-109, 1980.
- [Curtis 81] S. Curtis, T. Lake, G. Simpson, "A Magical Mystery Tour of the Motorola MC68000," unpublished, 1981.
- [Dang 77] L.-G. Dang, P. B. Ashkin, R. Yee, and M. O'Brein, "A CMOS/SOS 16-Bit Parallel μ CPT," *IEEE International Solid-State Circuits Conf.*, 134-135, 236, 1977.
- [Donath 79] W. E. Donath, "Placement and Average Interconnection Lengths of Computer Logic," *IEEE Trans. Circuits and Systems*, 272-277, 1979.
- [Flores 63] Ivan Flores, *The Logic of Computer Arithmetic*, Prentice-Hall, Englewood Cliffs, N. J., 1963.
- [Goel 82] P. Goel and B. C. Rosales, "PODEM-X: An Automatic Test Generation System for VLSI Logic Structures," *18th Design Automation Conference*, 260-268, Nashville, Tenn., 1982.
- [Goel 84] P. Goel, private communication, 1984.
- [Haviland 80] G. L. Haviland and A. A. Tuszynski, "A CORDIC Arithmetic Processor Chip," *IEEE J. Solid-State Circuits* SC-15, 1980.
- [Heller 78] W. R. Heller, W. F. Mikhail, W. E. Donath, "Prediction of Wiring Space Requirements for LSI," *Journal of Design Automation and Fault Tolerant Computing*, 117-144, 1978.
- [Hennessy 81] J. L. Hennessy, N. Jouppi, F. Baskett, and J. Gill, "MIPS: A VLSI Processor Architecture," *Proc. CMU Conf. VLSI Systems and Computations*, 337-348, Pittsburgh, Pa., 1981.
- [Hiatt 80] J. Hiatt, "Microprobing," *Proc. IEEE International Reliability Physics Symposium*, 116-120, 1980.
- [Hill 74] F. J. Hill and G. R. Peterson, *Introduction to Switching Theory and Logical Design*, Wiley, New York, 1974.

- [Hillis 81] W. D. Hillis, "The Connection Machine," AI memo no. 646, Massachusetts Institute of Technology, 1981.
- [Hwang 79] K. Hwang, *Computer Arithmetic, Principles, Architecture and Design*, Wiley, New York, 1979.
- [Johnson 84] W. N. Johnson, "A VLSI Superminicomputer CPU," *IEEE International Solid-State Circuits Conf.*, 174-175, San Francisco, Calif., 1984. (VAX V11)
- [Kaminker 81] A. Kaminker, L. Kohn, Y. Lavi, A. Menachem, and Z. Soha, "A 32-Bit Microprocessor with Virtual Memory Support," *IEEE J. Solid-State Circuits* SC-16: 548-557, 1981. (The NS16032)
- [Keyes 79] R. W. Keyes, "The Evolution of Digital Electronics Towards VLSI," *IEEE J. Solid-State Circuits* SC-14: 193-201, 1979.
- [Lutz 84] C. Lutz, S. Rabin, C. Seitz, and D. Speck, "Design of the Mosaic Element," *Proc. Advanced Research in VLSI*, 1-10, Cambridge, Mass., 1984.
- [Lyon 82] R. F. Lyon and M. P. Haerberli, "Designing and Testing the Optical Mouse," *VLSI Design* 3, no. 1, 20-30, 1982.
- [Magar 82] S. S. Magar, E. R. Caudel, and A. W. Leigh, "A Microcomputer with Digital Signal Processing Capability," *IEEE International Solid-State Circuits Conf.*, 32-33, 284-285, San Francisco, Calif., February 1982. (TMS 320)
- [Morse 78] S. P. Morse, W. B. Pohlman, and B. W. Ravenel, "The Intel 8086 Microprocessor: A 16-bit Evolution of the 8080," *IEEE Computer* 11: 18-27, June, 1978.
- [Murphy 81] B. T. Murphy, R. Edwards, L. C. Thomas, and J. J. Molinelli, "A CMOS 32b Single Chip Microprocessor," *IEEE International Solid-State Circuits Conf.*, 230-231, New York, N.Y., 1981. (The Bellmac 32)
- [NBS 83] Standard test structures and programs for certain technologies are available from the National Bureau of Standards and the Jet Propulsion Laboratory. Address inquiries to:
- | | | |
|--|-----|---|
| VLSI Test Chip Laboratory
Jet Propulsion Laboratory
4800 Oak Grove Drive
Pasadena, CA 91109 | and | National Bureau of Standards
Room B-310
Technology Building
Washington, DC 20234 |
|--|-----|---|
- [Obrebska 82] M. Obrebska, "Comparative Survey of Different Design Methodologies for Control Part of Microprocessors," Doctoral Thesis, 1982.
- [Obrebska 83] M. Obrebska, "Algorithm Transformations Improving Control Part Implementation," *IEEE International Conf. on Computer Design, VLSI in Computers*, 307-310, Port Chester, N.Y., 1983.
- [Patterson 82] D. A. Patterson and C. H. Séquin, "A VLSI RISC," *IEEE Computer* 15: 8-22, September 1982.
- [Richardson 81] W. S. Richardson, J. A. Bayliss, S. R. Colley, R. H. Kravitz, and G. A. McCormick, "The 32b Computer Instruction Decoding Unit," *IEEE International Solid-State Circuits Conf.*: 114-115, New York, N.Y., 1981. (The i432)
- [Roth 80] J. P. Roth, *Computer Logic, Testing, and Verification*, Computer Science Press, Potomac, Md., 1980. (The D-Algorithm)
- [Rowen 84] C. Rowen, S. A. Przbylski, N. P. Jouppi, T. Gross, J. Shott, and J. L. Hennessy, "A Pipelined 32b, NMOS Microprocessor," *IEEE International Solid-State Circuits Conf.*: 180-181, San Francisco, Calif., 1984 (Stanford MIPS)
- [Schumann 84] R. Schumann and W. Parker, "A 32b Bus Interface Chip," *IEEE International Solid-State Circuits Conf.*, 176-177, San Francisco, Calif., 1984.
- [Schuster 84] M. D. Schuster and R. E. Bryant, "Concurrent Fault Simulation of MOS Digital Circuits," *Proc. Advanced Research in VLSI*, 129-138, Cambridge, Mass., 1984.
- [Sherburne 82] R. W. Sherburne, Jr., M. G. H. Katevenis, D. A. Patterson, and C. H. Séquin, "Data-path Design for RISC," *Proc. Advanced Research in VLSI*, 53-62, Cambridge, Mass., 1982.
- [Sherburne 84] R. W. Sherburne, Jr., M. G. H. Katevenis, D. A. Patterson, and C. H. Séquin, "A 32b

- NMOS Microprocessor with a Large Register File." *IEEE International Solid-State Circuits Conf.*, 168-169. San Francisco, Calif., 1984. (Berkeley RISC)
- [Slager 83] J. Slager, L. Gindraux, T. Ho, G. Louie, and D. Vannier, "A 16b Microprocessor with On-Chip Memory Protection," *IEEE International Solid-State Circuits Conf.*, 24-25. New York. 1983. (iAPX 286)
- [Shima 79] M. Shima. "Demystifying Microprocessor Design." *IEEE Spectrum* 16, 22-30, July 1979.
- [Stritter 78] E. Stritter and H. L. Tredonick. "Microprogrammed Implementation of a Single Chip Microprocessor," *Proceedings of the 11th Annual Microprogramming Workshop*, 8-16, 1978. (MC68000)
- [Stritter 79] E. Stritter and T. Gunter. "A Microprocessor Architecture for a Changing World: The Motorola 68000," *IEEE Computer* 12, 43-52, Feb. 1979.
- [Sutton 83] J. A. Sutton, "The 80C86 CMOS Design from the nMOS Architecture." *VLSI Design* 4, no. 8, 56-63, 1983.

附 录

A. SPICE 技术*

- 输入文件用 `·END` 命令结尾，最后的回车可以省去。
- 为了确定一悬浮节点上的电压，人们把一个 $10\text{M}\Omega$ 的电阻接在该节点和地之间：`R100 0 10 M`。这种表示方式中的 M 只表示毫，不表示兆。兆欧要用 `MEG` 表示。
- 人们在输入卡中使用 `Tab` 键，这是不好的。同时，太长的行将被截断，因而不会显示错误信息。模型名、文件名和元件名的字符数目皆有限制。
- 人们常常忘记把 P 型器的 $\square$ 阱引出端连接到 V_{DD} （注意这是对本书所用 $\square$ 阱 CMOS 工艺而言）。
- 如果我们要模拟的电路，具有两个电压相同但数值未知的节点，则用 1H 电感把这两节点连接起来。
- 应当记住在晶体管的宽度或长度后面要写上单位“U”。在用更精巧的版本时，写 SPICE 的预处理程序时偶尔会输出 `12E-3U`，这应理解为 `12E-3`。
- 令源扩散区面积 $AS=1000\text{U}$ 。这意思是 $AS=1000\text{P}$ 吗？（注意，在 SPICE 中，后缀 U 表示微米 ($1\text{E}-6\text{M}$)，P 表示平方微米 ($1\text{E}-12\text{M}^2$)）。
- 在同一电路文件中，混合使用 Level 2 和 Level 3 MOS 模型时要留心，据报道可能会发生难以理解的结果。
- 如果直流分析不收敛，则提高模拟温度，重新进行直流模拟，并把这些数值用于 `NO-DESET` 指令中。这样处理后的节点，由 SPICE 建立的内矩阵是不同的。
- 利用一个电流控制电流源驱动一个线性电容，以便监视网络支路中的总电荷。可以用该技术监视 CMOS 电路的动态功耗，这将需要一些 $10\text{M}\Omega$ 的电阻。
- 人们把电路信息放在输入文件的第一行。这行被认为是标题信息。注意标题行是第一行，而不是第一非空白行。
- 如果在绘制直流转移函数图形时出现不收敛，那么就要缩小步长。在数字电路情况下，如果电路的增益高得不合理，通常会出现这种不收敛的问题，那么就要降低增益。
- 在瞬态分析中，在不同的地方串联 1Ω 电阻和并联 $1\text{G}\Omega$ 电阻，有助于解决“步长太小”的问题。通常，该问题是由电路中以很快时间常数开关的节点引起的。有时，在这些节点上附加电容将有助于进行模拟，而不会改变模拟结果。
- 当一个宽晶体管的宽度加倍时，电流增加不到二倍。设 $RSH=0$ 或规定 RD 和 RS 的值。源和漏电阻方块数的补缺值不为零。

B. 非线性电容方程

$$I \equiv -\frac{dQ}{dt}$$

* 为使意义明确，翻译时对原附录作了删改和补充。——译者注

$$Q = \int I dt$$

$$\Delta t = \int \frac{1}{I} dQ$$

$$C \equiv \frac{dQ}{dV}$$

$$I = C \frac{dV}{dt}$$

$$Q = \int C dV$$

$$\Delta t = \int \frac{C}{I} dV$$

C. 2 μ m nMOS 工艺范例

2 μ m nMOS 工艺范例采用砷源/漏注入和磷掺杂多晶硅。它有三种晶体管型式, 分别为注入增强型, 注入耗尽型和零阈值型。它被设计成以 $-3V V_{BB}$ 偏压工作。

2 μ m nMOS 工艺电学参量 (23 $^{\circ}C$ $V_{BB} = -3V$)

参 量	最 小 值	最 大 值	单 位	注 解
T_{ox}	270	330	$\text{\AA}$	栅氧化层厚度
C_{ox}	10.45	12.77	$\times 10^{-4} \text{pF}/\mu\text{m}^2$	栅氧化层电容
ΔL_{poly}	-0.125	0.125	μm	一侧
x_j	0.27	0.33	μm	源/漏结深
L_{diff}	0.19	0.23	μm	一侧
ΔW	0.6	0.9	μm	向内一侧
R_{poly}	15	40	$\Omega/\square$	多晶硅的薄层电阻
R_{diff}	20	50	$\Omega/\square$	n^+ 扩散区薄层电阻
C_{js}	0.2	0.3	$\times 10^{-4} \text{pF}/\mu\text{m}^2$	结区电容
C_{jp}	1.0	2.0	$\times 10^{-4} \text{pF}/\mu\text{m}^2$	结的周边电容
C_{poly}	0.5	0.7	$\times 10^{-4} \text{pF}/\mu\text{m}^2$	场区上的多晶硅电容
$C_{m, poly}$	0.82	1.10	$\times 10^{-4} \text{pF}/\mu\text{m}^2$	多晶硅上的金属电容
$C_{m, diff}$	0.82	1.1	$\times 10^{-4} \text{pF}/\mu\text{m}^2$	n^+ 扩散区上的金属电容
C_{mj}	0.31	0.42	$\times 10^{-4} \text{pF}/\mu\text{m}^2$	场区上金属的电容
LRC	4	10	$\Omega/\square$	多晶硅薄层/ n^+

2 μ m nMOS 工艺器件参量¹⁾ (23 $^{\circ}C$ $V_{BB} = -3V$)

按最小特征线宽 6 μ m 画版图

参量	增 强 型		零 阈 型		耗 尽 型		单 位
	最 小 值	最 大 值	最 小 值	最 大 值	最 小 值	最 大 值	
V_{T0}	0.6	1.0	0.2	0.6	-22	-1.7	V
μ	650	750	680	780	650	750	$\text{cm}^2/(\text{V} \cdot \text{s})$
K'	33.9	47.9	35.5	49.8	39.2	56.8	$\mu\text{A}/\text{V}^2$
N	2	5	1.5	4.0	20	50	$\times 10^{14}/\text{cm}^3$
γ	0.065	0.12	0.056	0.116	0.102	0.25	$\text{V}^{1/2}$
Theta	0.09	0.1	0.09	0.1	0.03	0.04	
K_{appa}	0.5	0.5	0.5	0.5	0.5	0.5	
Eta	0.22	0.22	0.22	0.22	0.22	0.22	
v_{max}	12	17	12	17	12	17	$\times 10^4 \text{m/s}$

1) 增强型管按 20/2 画, 零阈值晶体管按 20/3 画, 耗尽型管按 6/6 画。

2 μ m NMOS最坏条件模型文件

```

*****
* SPICE Models for 2 um NMOS
*****
* Worst-case slow parameters
*
.MODEL NENHS NMOS LEVEL=3 RSH=0 TOX=330E-10 LD=0.19E-6 XJ=0.27E-6
+ VMAX=13E4 ETA=0.25 KAPPA=0.5 NSUB=5E14 UO=650 THETA=0.1
+ VTO=0.946 CGSO=2.43E-10 CGDO=2.43E-10 CJ=6.9E-5 CJSW=3.3E-10
+ PB=0.7 MJ=0.5 MJSW=0.3 NFS=1E10
*
.MODEL NZEROS NMOS LEVEL=3 RSH=0 TOX=330E-10 LD=0.19E-6 XJ=0.27E-6
+ VMAX=13E4 ETA=0.25 KAPPA=0.5 NSUB=40E14 UO=650 THETA=0.1
+ VTO=0.526 CGSO=2.43E-10 CGDO=2.43E-10 CJ=6.9E-5 CJSW=3.3E-10
+ PB=0.7 MJ=0.5 MJSW=0.3 NFS=1E10
*
.MODEL NDEPS NMOS LEVEL=3 RSH=0 TOX=330E-10 LD=0.19E-6 XJ=0.27E-6
+ VMAX=13E4 ETA=0.25 KAPPA=0.5 NSUB=50E14 UO=650 THETA=0.04
+ VTO=-2.078 CGSO=2.43E-10 CGDO=2.43E-10 CJ=6.9E-5 CJSW=3.3E-10
+ PB=0.7 MJ=0.5 MJSW=0.3 NFS=1E10
*
* deltaLpoly (one side)=-0.125U
* deltaW (one side)=0.9U
*
*****
* Typical
*
.MODEL NENHT NMOS LEVEL=3 RSH=0 TOX=300E-10 LD=0.21E-6 XJ=0.3E-6
+ VMAX=15E4 ETA=0.18 KAPPA=0.5 NSUB=3.5E14 UO=700 THETA=0.095
+ VTO=0.781 CGSO=2.8E-10 CGDO=2.8E-10 CJ=5.75E-5 CJSW=2.48E-10
+ PB=0.7 MJ=0.5 MJSW=0.3 NFS=1E10
*
.MODEL NZEROT NMOS LEVEL=3 RSH=0 TOX=300E-10 LD=0.21E-6 XJ=0.3E-6
+ VMAX=15E4 ETA=0.18 KAPPA=0.5 NSUB=2.75E14 UO=730 THETA=0.095
+ VTO=0.354 CGSO=2.8E-10 CGDO=2.8E-10 CJ=5.75E-5 CJSW=2.48E-10
+ PB=0.7 MJ=0.5 MJSW=0.3 NFS=1E10
*
.MODEL NDEPT NMOS LEVEL=3 RSH=0 TOX=300E-10 LD=0.21E-6 XJ=0.3E-6
+ VMAX=15E4 ETA=0.18 KAPPA=0.5 NSUB=35E14 UO=700 THETA=0.035
+ VTO=-2.231 CGSO=2.8E-10 CGDO=2.8E-10 CJ=5.75E-5 CJSW=2.48E-10
+ PB=0.7 MJ=0.5 MJSW=0.3 NFS=1E10
*
* deltaLpoly(one side)=0U
* deltaW(one side)=0.75U
*
*****
* Fast
*
.MODEL NENHF NMOS LEVEL=3 RSH=0 TOX=270E-10 LD=0.23E-6 XJ=0.33E-6
+ VMAX=17E4 ETA=0.10 KAPPA=0.5 NSUB=2E14 UO=750 THETA=0.09
+ VTO=0.612 CGSO=3.4E-10 CGDO=3.4E-10 CJ=4.6E-5 CJSW=1.65E-10
+ PB=0.7 MJ=0.5 MJSW=0.3 NFS=1E10

```

```

*
.MODEL NZEROF NMOS LEVEL=3 RSH=0 TOX=270E-10 LD=0.23E-6 XJ=0.33E-6
+ VMAX=17E4 ETA=0.10 KAPPA=0.5 NSUB=1.5E14 UO=780 THETA=0.09
+ VTO=0.179 CGSO=3.4E-10 CGDO=3.4E-10 CJ=4.6E-5 CJSW=1.65E-10
+ PB=0.7 MJ=0.5 MISW=0.3 NFS=1E10
*
.MODEL NDEPF NMOS LEVEL=3 RSH=0 TOX=270E-10 LD=0.23E-6 XJ=0.33E-6
+ VMAX=17E4 ETA=0.10 KAPPA=0.5 NSUB=20E14 UO=750 THETA=0.03
+ VTO=-2.384 CGSO=3.4E-10 CGDO=3.4E-10 CJ=4.6E-5 CJSW=1.65E-10
+ PB=0.7 MJ=0.5 MISW=0.3 NFS=1E10
*
* deltaLpoly(one side)=0.125U
* deltaW(one side)=0.6U
*
*****
* Noise Margin
*
* (Assumes VIN near VDD, Wpu<Wpd and Lpu>Lpd)
*
.MODEL NENHN NMOS LEVEL=3 RSH=0 TOX=270E-10 LD=0.19E-6 XJ=0.27E-6
+ VMAX=15E4 ETA=0.25 KAPPA=0.5 NSUB=2E14 UO=670 THETA=0.1
+ VTO=0.577 CGSO=2.9E-10 CGDO=2.9E-10 CJ=5.75E-5 CJSW=2.48E-10
+ PB=0.7 MJ=0.5 MISW=0.3 NFS=1E10
*
.MODEL NZERON NMOS LEVEL=3 RSH=0 TOX=270E-10 LD=0.19E-6 XJ=0.27E-6
+ VMAX=15E4 ETA=0.18 KAPPA=0.5 NSUB=1.5E14 UO=730 THETA=0.095
+ VTO=0.365 CGSO=2.9E-10 CGDO=2.9E-10 CJ=5.75E-5 CJSW=2.48E-10
+ PB=0.7 MJ=0.5 MISW=0.3 NFS=1E10
*
.MODEL NDEPN NMOS LEVEL=3 RSH=0 TOX=270E-10 LD=0.19E-6 XJ=0.27E-6
+ VMAX=17E4 ETA=0.10 KAPPA=0.5 NSUB=20E14 UO=730 THETA=0.03
+ VTO=-2.392 CGSO=2.9E-10 CGDO=2.9E-10 CJ=5.75E-5 CJSW=2.48E-10
+ PB=0.7 MJ=0.5 MISW=0.3 NFS=1E10
*
* deltaLpoly(one side)=-0.125U
* deltaW(one side)=0.6U
*
*****
* Slow Fast
*
.MODEL NENHSF NMOS LEVEL=3 RSH=0 TOX=300E-10 LD=0.23E-6 XJ=0.33E-6
+ VMAX=17E4 ETA=0.10 KAPPA=0.5 NSUB=2E14 UO=730 THETA=0.1
+ VTO=0.614 CGSO=2.9E-10 CGDO=2.9E-10 CJ=5.75E-5 CJSW=2.64E-10
+ PB=0.7 MJ=0.5 MISW=0.3 NFS=1E10
*
.MODEL NZEROSF NMOS LEVEL=3 RSH=0 TOX=300E-10 LD=0.23E-6 XJ=0.33E-6
+ VMAX=15E4 ETA=0.18 KAPPA=0.5 NSUB=4E14 UO=700 THETA=0.095
+ VTO=0.353 CGSO=2.9E-10 CGDO=2.9E-10 CJ=5.75E-5 CJSW=2.64E-10
+ PB=0.7 MJ=0.5 MISW=0.3 NFS=1E10
*
.MODEL NDEPSF NMOS LEVEL=3 RSH=0 TOX=300E-10 LD=0.23E-6 XJ=0.33E-6
+ VMAX=12E4 ETA=0.25 KAPPA=0.5 NSUB=50E14 UO=670 THETA=0.03

```

+ VTO = -2 CGSO = 2.9E-10 CGDO = 2.9E-10 CJ = 5.75E-5 CISW = 2.64E-10
 + PB = 0.7 MJ = 0.5 MJSW = 0.3 NFS = 1E10

*
 * deltaLpoly(one side) = 0.125U
 * deltaW(one side) = 0.9U
 *

2 μm nMOS版图设计规则

层	规则	说 明	尺 寸
1.0		有源区掩模	
	1.1	最小宽度	3 μm
	1.2	最小间距	2 μm
2.0		增强注入掩模	
	2.1	围绕栅区的最小扩展	2 μm
	2.2	与无关栅区的最小距离	2 μm
3.0		耗尽注入掩模	
	3.1	围绕栅区的最小扩展	2 μm
	3.2	与无关栅区的最小距离	2 μm
4.0		隐埋接触掩模	
	4.1	围绕接触区的最小扩展	1 μm
	4.2	接触区与无关多晶硅或扩散区的最小距离	2 μm
5.0		多晶硅掩模	
	5.1	最小宽度	2 μm
	5.2	最小间距	2 μm
	5.3	与无关扩散区的最小距离	1 μm
	5.4	超出栅区的最小尺寸	2 μm
	5.5	为形成隐埋接触, 多晶硅与有源区掩模的最小重叠面积	2 μm × 3 μm
	5.6	为形成隐埋接触, 在场区方向上有源区掩模超出多晶硅的最小尺寸	1 μm
6.0	5.7	有源区超出栅区的最小尺寸	3 μm
		金属接触掩模 ¹⁾	
	6.1	最小接触面积	2 μm × 2 μm
	6.2	有源区掩模对n ⁺ 接触的最小扩展	2 μm
	6.3	n ⁺ 接触与多晶硅的最小距离	2 μm
	6.4	多晶硅掩模对多晶硅接触的最小扩展	2 μm
7.0	6.5	多晶硅接触与有源栅区的最小间距	2 μm
		金属掩模 ²⁾	
	7.1	最小宽度	4 μm
	7.2	最小间距	2 μm
	7.3	金属超出接触的最小尺寸	1 μm
8.0		钝化掩模	
	8.1	最小压焊块窗口尺寸	100 μm × 100 μm
	8.2	最小压焊块-压焊块间距	75 μm
		金属超出压焊块的最小尺寸	4 μm

1) 接触的最大电流密度是 0.2mA/μm (沿周边)。

2) 金属线的最大电流密度是 0.7mA/μm (沿宽度)。

D. 2 μ m CMOS 工艺范例

2 μ m CMOS 工艺范例采用增强型 P 和 n 沟器件, 并使用 n 阱技术。采用低阻外延层以提供良好的自锁保护。通常仅在内电路的阱中需要接点。n 沟器件用神源/漏注入, 而 P 沟器件则用硼。n 沟和 P 沟器件都采用场注入。

2 μ m CMOS 工艺电学参量 (n 阱, 23 $^{\circ}$ C)

参 量	最 小 值	最 大 值	单 位	注 释
T_{ox}	225	275	$\text{\AA}$	
C_{ox}	12.55	15.34	$\times 10^{-4} \text{pF}/\mu\text{m}^2$	
ΔL_{poly}	-0.125	0.125	μm	向内一侧
ΔW	0.6	0.9	μm	向内一侧
$C_{m\ poly}$	0.75	1.03	$\times 10^{-4} \text{pF}/\mu\text{m}^2$	加上边缘电容
$C_{poly\ t}$	0.47	0.53	$\times 10^{-4} \text{pF}/\mu\text{m}^2$	加上边缘电容
$C_{m\ diff}$	0.75	0.93	$\times 10^{-4} \text{pF}/\mu\text{m}^2$	加上边缘电容
$C_{m\ t}$	0.29	0.35	$\times 10^{-4} \text{pF}/\mu\text{m}^2$	加上边缘电容
R_{m/p^+}		5	Ω	2 $\times$ 2 接触
R_{m/n^+}		10	Ω	2 $\times$ 2 接触
$R_{m/poly}$		2	Ω	2 $\times$ 2 接触

2 μ m CMOS 工艺器件参量¹⁾ (n 阱, 23 $^{\circ}$ C)

按最小特征线宽 6 μ m 画版图

参 量	n 沟		P 沟		单 位
	最 小 值	最 大 值	最 小 值	最 大 值	
V_{TO}	0.6	1.0	-1.0	-0.6	V
μ	550	650	180	220	$\text{cm}^2/(\text{V}\cdot\text{s})$
K'	35	50	11	17	$\mu\text{mho}/\text{V}$
N	3	4	5	7	$\times 10^{15} \text{cm}^{-3}$
γ	0.16	0.3	0.3	0.38	$\sqrt{V}$
L_{diff}	0.1	0.15	0.3	0.4	μm
X_j	0.14	0.21	0.42	0.6	μm
R_{poly}	20	50	20	60	$\Omega/\square$
C_{js}	1.0	1.6	6.0	7.7	$\times 10^{-4} \text{pF}/\mu\text{m}^2$
C_{jp}	1.25	1.8	3.75	5.4	$\times 10^{-4} \text{pF}/\mu\text{m}^2$
R_{diff}	20	50	80	120	$\Omega/\square$

1) 晶体管尺寸按 15/2 画。

2 μ m CMOS 曼坏条件模型文件

```

*****
* 2 um n-Well Iso CMOS
*****
* Slow Slow
*
.MODEL NSS NMOS LEVEL=3 RSH=0 TOX=275E-10 LD=0.1E-6 XJ=0.14E-6
+ CJ=1.6E-4 CJSW=1.8E-10 UO=550 VTO=1.022 CGSO=1.3E-10
+ CGDO=1.3E-10 NSUB=4E15 NFS=1E10

```

```

+ VMAX=12E4 PB=0.7 MJ=0.5 MJSW=0.3 THETA=0.06 KAPPA=0.4 ETA=0.14
.MODEL PSS PMOS LEVEL=3 RSH=0 TOX=275E-10 LD=0.3E-6 XJ=0.42E-6
+ CJ=7.7E-4 CJSW=5.4E-10 UO=180 VTO=-1.046 CGSO=4E-10
+ CGDO=4E-10 TPG=-1 NSUB=7E15 NFS=1E10
+ VMAX=12E4 PB=0.7 MJ=0.5 MJSW=0.3 ETA=0.06 THETA=0.03 KAPPA=0.4
*
* deltaLpoly=-0.125um deltaW=0.9um (one sided inward)
*
*****
* Fast p-type Slow n-type
*
.MODEL NFS NMOS LEVEL=3 RSH=0 TOX=250E-10 LD=0.1E-6 XJ=0.14E-6
+ CJ=1.6E-4 CJSW=1.5E-10 UO=550 VTO=1.03 CGSO=1.33E-10
+ CGDO=1.33E-10 NSUB=4E15 THETA=0.06 KAPPA=0.4 ETA=0.14
+ VMAX=12E4 PB=0.7 MJ=0.5 MJSW=0.3 NFS=1E10
.MODEL PFS PMOS LEVEL=3 RSH=0 TOX=250E-10 LD=0.4E-6 XJ=0.6E-6
+ CJ=7E-4 CJSW=4.5E-10 UO=220 VTO=-.66 CGSO=5.5E-10
+ CGDO=5.5E-10 TPG=-1 NSUB=5E15 ETA=0.06 THETA=0.03 KAPPA=0.4
+ VMAX=17E4 PB=0.7 MJ=0.5 MJSW=0.8 NFS=1E10
*
* deltaLpoly=0um deltaWp=0.7um deltaWn=0.8um (one sided inward)
*
*****
* Fast p-type Fast n-type
*
.MODEL NFF NMOS LEVEL=3 RSH=0 TOX=225E-10 LD=0.15E-6 XJ=0.21E-6
+ CJ=1.0E-4 CJSW=1.25E-10 UO=650 VTO=0.628 CGSO=2.3E-10
+ CGDO=2.3E-10 NSUB=3E15 THETA=0.06 KAPPA=0.4 ETA=0.14
+ VMAX=17E4 PB=0.7 MJ=0.5 MJSW=0.3 NFS=1E10
.MODEL PFF PMOS LEVEL=3 RSH=0 TOX=225E-10 LD=0.4E-6 XJ=0.6E-6
+ CJ=6E-4 CJSW=3.75E-10 UO=220 VTO=-.668 CGSO=6.2E-10
+ CGDO=6.2E-10 TPG=-1 NSUB=5E15 ETA=0.06 THETA=0.03 KAPPA=0.4
+ VMAX=17E4 PB=0.7 MJ=0.5 MJSW=0.3 NFS=1E10
*
* deltaLpoly=0.125um deltaW=0.6um (one sided inward)
*
*****
* Slow p-type Fast n-type
*
.MODEL NSF NMOS LEVEL=3 RSH=0 TOX=250E-10 LD=0.15E-6 XJ=0.21E-6
+ CJ=1.0E-4 CJSW=1.5E-10 UO=650 VTO=0.826 CGSO=2E-10
+ CGDO=2E-10 NSUB=3E15 THETA=0.06 KAPPA=0.4 ETA=0.14
+ VMAX=17E4 PB=0.7 MJ=0.5 MJSW=0.3 NFS=1E10
.MODEL PSF PMOS LEVEL=3 RSH=0 TOX=250E-10 LD=0.3E-6 XJ=0.42E-6
+ CJ=7E-4 CJSW=4.5E-10 UO=180 VTO=-1.049 CGSO=4.2E-10
+ CGDO=4.2E-10 TPG=-1 NSUB=7E15 ETA=0.06 THETA=0.03 KAPPA=0.4
+ VMAX=12E4 PB=0.7 MJ=0.5 MJSW=0.3 NFS=1E10
*
* deltaLpoly=0um deltaWp=0.8um deltaWn=0.7um (one sided inward)
*
*****

```

2 μm CMOS版图设计规则

层	规则	说 明	尺 寸
1.0		n阱掩模	
	1.1	最小宽度	2 μm
	1.2	相同电位阱之间的最小间距	2 μm
2.0	1.3	不同电位阱之间的最小间距	2 μm
		有源区掩模	
	2.1	最小宽度	3 μm
	2.2	最小间距	2 μm
	2.3	阱内有源区到阱边的距离	2 μm
3.0	2.4	阱外有源区到阱边的距离	6 μm
		多晶硅掩模	
	3.1	最小宽度	2 μm
	3.2	最小间距	2 μm
	3.3	与无源区的最小间距	1 μm
	3.4	超出有源区的最小尺寸	2 μm
	3.5	有源区超出栅极的最小尺寸	3 μm
4.0		p ⁺ 掩模 (n ⁺ 掩模是p ⁺ 的反版)	
	4.1	最小宽度	2 μm
	4.2	最小间距	2 μm
	4.3	超出p ⁺ 有源区的最小尺寸	4 μm
	4.4	到n ⁺ 有源区的最小间距	4 μm
	4.5	超出pMOS晶体管的最小尺寸	4 μm
	4.6	到nMOS晶体管的最小尺寸	4 μm
	4.7	p ⁺ /n ⁺ 短路接触	特定
5.0		接触掩模	
	5.1	最小接触面积	2 $\mu\text{m} \times 2 \mu\text{m}$
	5.2	除p ⁺ /n ⁺ 短路接触外的最大接触面积	2 $\mu\text{m} \times 2 \mu\text{m}$
	5.3	p ⁺ /n ⁺ 短路接触尺寸	2 $\mu\text{m} \times 6 \mu\text{m}$
	5.4	与有源区周边最小距离	2 μm
	5.5	到栅区的最小间距	2 μm
6.0	5.6	与多晶硅周边的最小距离	2 μm
		金属	
	6.1	最小宽度	4 μm
	6.2	最小间距	2 μm
7.0	6.3	超出接触的最小尺寸	1 μm
		钝化掩模	
	7.1	最小压焊宽尺寸	100 $\mu\text{m} \times 100 \mu\text{m}$
	7.2	最小压焊块-压焊块间距	75 μm
8.0	7.3	金属超出压焊块的最小尺寸	4 μm
		p ⁺ /n ⁺ 特殊短路接触。该接触由一个位于10 $\times$ 6 μm 有源区中央的6 $\times$ 2 μm 接触孔构成。p ⁺ 掩模覆盖该区的一半 (5 $\times$ 6 μm)，就把接触在中间分开，形成一个3 $\times$ 2 n ⁺ 和一个3 $\times$ 2 p ⁺ 的接触	
9.0		阱接触掩模	
	9.1	接触之间的最小距离	100 μm
	9.2	接触之间最小方块数	5
10.0		保护环等	
	10.1	必须用和保护I/O结构相同的方式保护所有内部自举节点	
	10.2	见 5.7 节	

E. nMOS和CMOS工艺的两项先进备选工艺

可以单独或一起选用下列两项备选工艺（增加一些费用）：

LRP 低阻多晶。利用硅化物， R_{pol} ，变为 2 到 4 $\Omega/\square$ 。

M2 第二层金属，设计规则是：

M2 接触孔（氧化层厚度 1 μm ）	
到 M_1 接触的最小间距	2 μm
超出 M_1 接触窗口的最小尺寸	2 μm
接触孔尺寸	2 $\mu\text{m} \times 2\mu\text{m}$
M2（金属厚度 1 μm ）	
最小间距	3 μm
最小宽度	6 μm
超出接触窗口的最小距离	2 μm

F. 有用的常数和换算

通用常数

q	$1.602 \times 10^{-19} \text{ C}$
ϵ_0	$8.854 \times 10^{-12} \text{ F/m}$
kT/q	25.86mV
kT	0.02586eV
μ_0	$4\pi \times 10^{-7} \text{ H/m}$
c	$2.998 \times 10^8 \text{ m/s}$
m_e	$9.11 \times 10^{-31} \text{ kg}$
h	$6.625 \times 10^{-34} \text{ J}\cdot\text{s}$
k	$1.38 \times 10^{-23} \text{ J/K}$

硅常数

（全部数据皆在 300K = 27°C = 80.6°F 温度下测得）

$\kappa_{\text{thermal, SiO}_2}$	0.014W/(cm·°C)	（热导）
$\kappa_{\text{thermal, Si}}$	1.412W/(cm·°C)	（热导）
κ_{SiO_2}	3.9	
κ_{Si}	11.7	
$\kappa_{\text{Si}_3\text{N}_4}$	7.5	
ϵ_{SiO_2}	$3.45 \times 10^{-11} \text{ F/m}$	
ϵ_{Si}	$1.04 \times 10^{-10} \text{ F/m}$	
$\epsilon_{\text{Si}_3\text{N}_4}$	$6.64 \times 10^{-11} \text{ F/m}$	
E_g	1.124eV (1.170eV, 0K)	

μ_n	$1350\text{cm}^2(\text{V}\cdot\text{s})$	
μ_p	$480\text{cm}^2/(\text{V}\cdot\text{s})$	
$E_{\text{breakdown, SiO}_2}$	$3 \times 10^7 \text{ V/m}$	(击穿电场)
$E_{\text{breakdown, SiO}_2}$	$6 \times 10^8 \text{ V/m}$	(击穿电场)
χ_{Si}	4.15eV	
n_i	$1.45 \times 10^{10} \text{ cm}^{-3}$	
N_c	$2.8 \times 10^{19} \text{ cm}^{-3}$	
N_v	$1.04 \times 10^{19} \text{ cm}^{-3}$	

单位换算

$$1\text{mil}=25.4\mu\text{m}=10^{-3}\text{in}$$

$$1\text{cm}=10^4\mu\text{m}=10^8\text{\AA}$$

$$1\text{eV}=1.602 \times 10^{-19} \text{ J}$$

$$0^\circ \text{C}=273 \text{K}$$

索引

二 画

二进制计数器 Binary counter, 31
二氧化硅 Silicon dioxide, 121

三 画

工艺范例 Example processes, 10, 331—338
工艺参量 Process parameters, 10
工艺测试结构 Process test structures, 322
小信号增益 Small signal gain, 159
小结 Summary
CMOS工艺~ CMOS process, 132
nMOS工艺~ nMOS process, 129
二阶效应~ second-order effects, 91
电流方程~ current equations, 64
通用MOS符号~ generic MOS notation, 27
阈值电压效应~ threshold voltage effects, 91
阈值方程~ threshold equations, 71
晶体管符号~ transistor notation, 8
最坏条件功耗~ worst-case power, 176
最坏条件速度~ worst-case speed, 176
最坏条件噪声~ worst-case noise, 179
上拉 Pullup, 11
下拉 Pulldown, 11
“与” AND, 321
“与”阵 AND plane, 35, 284, 314, 316

四 画

中继器 Repeater, 190
水平控制 Horizontal control, 317
分层结构 Hierarchy, 1, 313, 319
化学汽相淀积 Chemical vapor deposition, 123, 128
双端读出 Double-ended sensing, 199, 292
双稳态 Bistability, 202, 204
计数器 Counter, 30
区熔提纯 Float zone refining, 117
反转电压 V_{inv} Inverse voltage V_{inv} , 155, 164
反型 Inversion, 59, 83
反转阱 Retrograde wells, 118
反相器 Inverter, 11
反射规则 Reflection rules, 133
介质隔离 Dielectric isolation, 38

开关能量 Energy, switching, 165

五 画

加法器 Adder, 12—14, 21, 37, 302, 308—309, 313
母版 Reticles, 119
生产测量 Manufacturing test, 323
冯·诺伊曼计算机 Von Neumann computer, 40
边沿速率 Edge rate, 197, 207, 237, 253
外延层 Epitaxial layer, 89, 118
鸟嘴 Bird beak, 72, 85, 121
对准符号 Alignment marks, 120
对称版图设计 Symmetric layout, 237
对接接触 Contact, butting, 137
对称性 Duality, 21, 23, 250
平面布置 Floor plan, 227
平带电压 Flat-band voltage, 53, 57—58, 82, 84
功函数 Work function, 57
功耗 Power dissipation, 107, 173, 178, 187, 188, 191
功耗延迟乘积 Power delay product, 157, 158, 192
电源分配 Power distribution, 231
电子束 Electron beam(*), 119, 120
电平转换 Level conversion, 215
电迁移 Electromigration, 93
电阻率 Resistivity, 68
电场 $\mathcal{E}$ Electric field $\mathcal{E}$, 4, 52, 65, 75, 111
电阻耦合噪声 Resistive coupling noise, 169
电流镜 Current mirror, 166, 208
电容 Capacitance
MOS~ MOS, 62, 68
扩散~ diffusion, 96, 160
互连~ interconnect, 187, 188
平行板~ parallel plate, 94
本征~ intrinsic, 73, 156
边缘~ edge, 94, 233
边缘~ fringing, 111
复叠~ overlap, 9, 73, 128
侧壁~ side-wall, 94, 96
固有~ self, 185
固有负载~ self-loading, 160
结~ junction, 52
氧化物~ oxide, 52
米勒~ Miller, 96, 157, 205
晶体管~ transistor, 69

输入~ input, 156—157, 160—162, 185, 192
 耦合~ coupling, 95, 232, 278
 电容控制 Capacitive steering, 212
 电荷分配 Charge sharing, 28, 36, 199, 250, 290
 电荷分配放大器 Charge sharing amplifier, 200
 电感 Inductance, 269
 电源噪声 Power supply noise, 232
 可测性设计 Design for testability, 324

六 画

设计规则 Design rule(s), 10, 132—144
 ~推导 derivation, 138
 关键通道 Critical path, 188, 244, 296, 297, 321
 共源共栅电路 Cascode, 199
 并发操作 Concurrency, 313
 先行 Look-ahead, 38
 先行进位 Carry look-ahead, 308
 有比CMOS Ratioed CMOS, 166, 181
 有限状态机 Finite-state machine, 40, 313
 有源区 Active area(s), 3, 72, 85, 127, 129, 132, 136
 异步事件 Asynchronous event, 264, 265
 异步系统 Asynchronous systems, 261
 导电因子 K' Conductance factor K' , 64, 67
 导带 Conduction band, 48, 49, 92
 价带 Valence band, 48, 49, 92
 自对准接触 Self-aligned contact, 129
 自定时电路 Self-timed circuitry, 261—264
 自定时信号 Self-timed signaling, 243
 自举 Bootstrap, 195—199, 256, 277—278, 292
 自举逻辑门 Bootstrapped logic gates, 16, 23
 自举缓冲器 Bootstrap buffers, 282
 自锁 Latchup, 85—87
 阵列结构 Array structure, 231
 阱接点 Well plug, 137
 迁移率 Mobility μ , 4, 60, 74—76, 91, 92, 175
 过驱动 Overdrive 191, 194
 扩散 Diffusion, 122
 扫描设计 Scan design, 325
 场注入 Field implant, 72, 73, 85, 126, 129
 场氧化 Field oxidation, 72, 73
 场氧化物 Field oxide, 129
 场阈值 Field threshold, 73
 优化 Optimization, 117, 185—188, 237, 282, 284, 299
 优先权编码器 Priority encoder, 306
 传输晶体管 Pass transistor, 11
 ~逻辑 logic, 11, 14, 22, 149
 伪随机顺序计数器 Pseudorandom sequence counter, 31

吸取(杂质) Gettering, 128
 次开启电导 Subthreshold conduction, 79
 次开启优值 α Subthreshold figure of merit alpha, 80
 次开启漏电 Subthreshold leakage, 175
 光复印 Photolithography, 118—120
 光致抗蚀剂 Photoresist, 120, 126, 133, 134
 多米诺逻辑 Domino logic, 248—250
 多层抗蚀剂 Multilayer resist, 133
 多晶 Poly, 2
 多晶硅电阻器 Poly resistors, 287
 多路转接器 Multiplexer, 316
 多端口寄存器单元 Multiported register cell, 36
 再定时 Retiming, 295, 296
 负载线 Load line, 149
 亚稳的 Metastable, 265
 压焊块驱动晶体管 Pad drive transistors, 233
 收缩阵列 Systolic arrays, 292—299
 同步 Synchronization, 25, 205, 264
 同步器 Synchronizer, 264—268, 272
 同步系统 Synchronous systems, 261
 成品率模型 Yield model, 141
 动态RAM Dynamic RAM, 289—292
 动态节点 Dynamic node(s), 171, 232, 252
 动态范围 Dynamic range, 209
 动态存储 Dynamic storage, 152, 168, 203
 动态锁存器 Dynamic latch, 203

七 画

位译码器 Bit decoder, 306
 体效应因子 γ Body factor γ , 63, 68, 85
 体效应 Body effect, 158, 164
 体效应参数 γ_s Body effect parameter γ_s , 83
 译码器 Decoder, 33, 34, 276
 ~树 tree, 276—277
 离子注入编程~ implant programmed, 276—277
 投影光刻机 Projection aligner, 118
 沟道 Channel, 6, 9, 64, 84, 85
 沟道长度调制 Channel length modulation, 76, 77, 164
 沟道效应 Channeling, 122
 沟槽隔离 Trench isolation, 88
 进位链 Carry chain, 306, 309
 迟滞 Hysteresis, 207
 形状因子 S Shape factor S , 9, 153, 190
 间距 Pitch, 230, 277, 311
 判优器 Arbiter, 264
 吞吐量 Throughput, 296—298
 时钟发生器 Clock generator(s), 256, 260, 268

时钟定时 Clocking, 243,
时钟, 非重叠 Clocks, nonoverlapping, 26, 27
时钟歪斜 Clock skew, 253, 254, 268, 269
时钟差分放大器 Clocked differential amplifier
210—212
时钟缓冲器 Clocked buffer(s), 194, 237

八 画

侧壁电容 Side wall capacitance, 73
刻蚀 Etching, 134
干法~ dry, 121
反应离子~ reactive ion, 121
各向异性~ anisotropic, 121
择优~ preferential, 121
湿法~ wet, 120
环形晶体管 Doughnut transistor, 235
环路计数器 Loop counter, 320
环路增益 Loop gain, 206, 287
转移 Branching, 316
传输函数 Transfer function, 149, 150, 153, 154,
156, 161, 164, 165, 207, 217
线性区 Linear region, 61
固有电容 Self-capacitance, 199
固定电荷 Fixed charge, 58
衬底偏置 V_{BB} Substrate bias V_{BB} , 218, 221—227
饱和 Saturation, 64, 67, 155, 164, 166
饱和比 X Saturation ratio X , 65, 70
规则性因子 Regularity factor, 32
规则结构 Regular structures, 2, 32—39, 42, 188,
237, 276—300
规则逻辑结构 Regular logic structures, 32
参考电压 Reference voltage, 202
受主 Acceptor, 49
变容二极管 Varactor, 56
定时容限 Timing margin, 254
定时信号发生器 Timing generator, 208
单管RAM单元 1T1 RAM cell, 289—292
单端读出 Single-ended sensing, 199, 292
势垒 Potential barrier, 6, 7, 63
垂直控制 Vertical control, 317
“或”阵 OR plane, 35, 282
非饱和 Nonsaturation, 64, 67, 164
非重叠时钟 Nonoverlapping clocks, 244, 256

九 画

泄放器件 Bleeder device(s), 167, 251, 285, 287
差分检测 Differential detection, 199
差分放大器 Differential amplifier
时钟~ clocked, 210—212
静态~ static, 208—210

差分电荷放大器 Differential charge amplifier,
212—214
差分读出放大器 Differential sense amplifier,
285

信号强度 Signal strength, 150
保留进位加法器 Carry save adder, 38, 313
按内容访问存储器 CAM, 300
栅氧化物 Gate oxide, 132
相平面分析技术 Phase plane technique, 205
测试 Testing, 124, 322—325
退火 Annealing, 122
选择器 Selector, 11, 13, 22
统计波动 Statistical variations, 72
统计效果 Statistical effects, 173
突变结 Step junction, 51
施主 Donor, 49
施密特触发器 Schmitt trigger, 202, 206—207, 222,
237
封装(管壳) Packaging, 107, 124
故障排除 Debug, 322—323
故障覆盖面 Fault coverage, 324
穿透 Punch-through, 79
穿接 Crossunder, 231
界面俘获电荷 Interface trapped charge, 58
费米势 Fermi potential, 49, 57
费米能级 Fermi level, 48
重叠时钟 Overlapping clocks, 258—260
版图设计 Layout, 3, 17, 19, 24, 142, 143, 207, 226,
227—237, 254, 281

十 画

流水作业 Pipelining, 320
流水线 Pipeline, 248, 293, 298
高电容单元 High-C cell, 292
高速缓冲存储器 Cache memory, 321
容错 Fault tolerance, 142
窄沟效应 Narrow channel effects, 72, 78, 85
读出放大器 Sense amplifier, 200, 208, 214, 231,
262, 291, 311
速度 v Velocity v , 4, 60, 75, 91
速度饱和 Velocity saturation, 74, 75, 76
积累 Accumulation, 53, 59
特征分析 Signature analysis, 325
耗尽 Depletion, 53
耗尽区 Depletion region, 3, 7, 51—56, 59, 63, 77,
78, 79, 83, 84
耗尽负载反相器 Depletion-load inverter, 17, 157
耗尽负载nMOS Depletion-load nMOS, 17, 23, 153
耗尽型晶体管 Depletion-mode transistor, 7, 8,
17, 83
热电子效应 Hot electron effect, 76, 87

热阻 Thermal resistance, 108
 扇入 Fan-in, 199
 扇出 Fan-out, 185, 188
 能带图 Energy band diagrams, 50
 禁带宽度 Band gap energy, 49
 竞争条件 Race condition, 245
 弱反型 Weak inversion, 80
 乘法器阵列 Multiplier array, 36—39
 离子注入 Ion implantation, 72, 81, 122, 127, 136
 缺陷密度 Defect density, 141
 氧化层俘获电荷 Oxide trapped charge, 58
 剥离 Lift off, 122
 预充电 Precharging, 32, 35, 152, 164, 237, 245—250, 255, 260, 278, 279—281, 308
 预取指流水线 Prefetch pipeline, 320

十 一 画

假信号 Glitch, 196
 闪烁预充时钟 Glitch precharge clock, 282
 深耗尽 Deep depletion, 56
 控制字 Control word, 314
 控制存储器 Control store, 40
 控制结构 Control structure, 313
 控制通路 Control path, 40, 305
 接触孔 Contact hole, 128
 接触电阻 Contact resistance, 92—93
 探针测试 Probing, 123
 掩模 Mask, 118, 125
 掩模对准 Mask alignment, 118, 120, 135—136, 138
 液桶式移位器 Barrel shifter, 307
 隐埋层 Buried layer, 118
 隐埋接触 Buried contact, 127, 130, 131, 136
 焊丝 Bonding wire, 124, 269
 焊丝键合 Wire bonding, 124
 移位寄存器 Shift register, 25, 26, 33, 247, 288
 移位器 Shifter, 306
 寄存器堆 Register file, 40, 227, 270, 306, 310
 逻辑功能发生器 Logic function generator, 306
 逻辑最小化 Logic minimization, 282
 硅片步进器 Wafer stepper, 120
 硅化物 Silicides, 93
 阈值电压 Threshold voltage V_T , 4, 16, 59, 63, 78, 84, 91, 153, 174
 虚地 Virtual ground, 277, 279, 281
 虚地ROM Virtual ground ROM, 280
 雪崩击穿 Avalanche breakdown, 52
 曼彻斯特进位链 Manchester carry chain, 250, 308
 毫微码 Nanocode, 317, 318

十 二 画

渡越时间 t_t , Transit time t_t , 4

溅射 Sputtering, 123
 缓冲氧化物 Buffer oxide, 129, 132
 缓冲器 Buffer, 189, 237
 短沟效应 Short channel effects, 76—79, 150, 232
 等待时间 Latency, 296—299
 锁存器 Latch, 25
 锁相 Phase-lock, 267
 超缓冲器 Superbuffer, 18—19, 24, 95, 191—194
 最坏条件文件 Worst-case files, 10
 最坏条件设计 Worst-case design, 172—181
 最坏条件参量 Worst-case parameters, 172—176, 179
 最坏条件模型 Worst-case models, 72
 最坏条件噪声容限 Worst-case noise margin, 179
 最坏条件噪声参量 Worst-case noise parameters, 163

十 三 画

源跟随器 Source follower, 216
 微处理器 Microprocessor, 228, 305
 微处理器系统定时 Microprocessor system timing, 269
 微码 Microcode, 317
 微程序 Microprogram, 40
 微程序设计 Microprogramming, 314
 输入编码 Input encoding, 284
 输出电导 g_D , Output conductance g_D , 155
 跨导 g_m , Transconductance g_m , 155
 数字滤波器 Digital filter, 295
 数据总线(母线) Data bus, 311
 数据通路 Data path, 40, 227, 229, 262, 306
 触发点 Trigger point, 206, 207
 触发器 Flip-flop, 202
 零阈值晶体管 Zero-threshold transistor, 8, 20

十 四 画

模式匹配 Pattern matching, 293
 模块性 Modularity, 229
 耦合 Coupling
 电容性~ capacitive, 168
 电阻性~ resistive, 169
 耦合电容 Coupling capacitance, 195
 静电放电 Electrostatic discharge, 87, 217—221
 静态RAM Static RAM, 286
 静态差分放大器 Static differential amplifier, 208—210
 静态锁存器 Static latch, 203—206
 截止 Cutoff, 64, 67
 算术逻辑运算部件 ALU (arithmetic logic unit), 40, 227, 270, 306, 310

漏感应势垒降低 Drain-induced barrier lowering, 79, 153, 164, 250

十五画

sneak path(s), 13, 279
增强型负载nMOS Enhancement-load nMOS, 11
增强型晶体管 Enhancement-mode transistor, 3, 7, 8
增量寄存器 Incrementing register, 310
横向扩散 Lateral diffusion, 73

十六画

噪声 Noise, 25, 232, 250, 267, 278, 287
噪声性能 Noise performance, 168—172
噪声相干性 Noise coherence, 169
薄层电阻 Sheet resistance, 92
薄膜 Thin film, 122

十七画

瞬态响应 Transient response, 157

其它

CMOS功耗 CMOS power dissipation, 165
CMOS静态逻辑 CMOS static logic, 20

Czochralski方法 Czochralski method, 117
FF模型 FF model, 175
FS模型 FS model, 176
p-n二极管 p-n diode, 50
PLA (可编程逻辑阵列) PLA (Programmable logic array), 33, 35, 40, 41, 199, 231, 233, 279—284, 306, 314, 315, 316, 321
PLA折叠 PLA folding, 283
pMOS上拉管 pMOS pullup, 166
RAM (随机存取存储器) RAM (Random-access memory), 33, 206, 282, 284—292
RAM单元 RAM cell, 284
RC延迟线 RC delay line, 190, 233, 291
RC梯形网络 RC ladder networks, 97
ROM (只读存储器) ROM (read-only memory), 32, 34—35, 40, 41, 99, 170, 214, 231, 233, 279—284, 306, 314, 316, 316, 319, 321
SLA (存储逻辑阵列) SLA (storage logic array), 284
SPICE技术 techniques 330
SS模型 SS model, 176
TTL电平转换 TTL level conversion, 215
 V_{DD} 反跳 V_{DD} bounce, 278
XOR门 XOR gate, 22, 180, 306, 319
 α -粒子 Alpha particles, 170
 β -比 β -ratio, 153, 160