

[19] 中华人民共和国国家知识产权局

[51] Int. Cl⁷

G11C 11/00

G11C 5/00

[12] 发明专利申请公开说明书

[21] 申请号 00805604.8

[43] 公开日 2002 年 4 月 17 日

[11] 公开号 CN 1345448A

[22] 申请日 2000.2.25 [21] 申请号 00805604.8

[30] 优先权

[32] 1999.3.26 [33] US [31] 09/277,347

[86] 国际申请 PCT/US00/04898 2000.2.25

[87] 国际公布 WO00/58969 英 2000.10.5

[85] 进入国家阶段日期 2001.9.26

[71] 申请人 硅芯片公司

地址 美国加利福尼亚州

[72] 发明人 D·K·刘 王鼎华

[74] 专利代理机构 中国专利代理(香港)有限公司

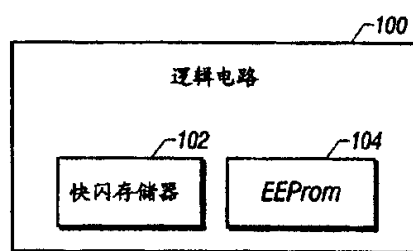
代理人 程天正 陈景峻

权利要求书 2 页 说明书 15 页 附图页数 8 页

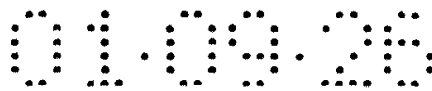
[54] 发明名称 带有嵌入的快闪和 EEPROM 存储器的器件

[57] 摘要

集成电路晶片(100)包括第一部分,其上包括逻辑电路。晶片的第二部分(104)包括 EEPROM 存储器,以及第三部分(102)包括 FLASH 存储器。

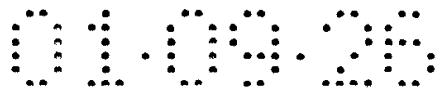


ISSN 1008-4274



权 利 要 求 书

1. 集成电路晶片包括：
在所述晶片上的 EEPROM 存储器；以及
在所述晶片上用与所述 EEPROM 存储器基本相同的处理步骤形成的
5 FLASH 存储器。
2. 权利要求 1 的晶片，还包括用与所述 EEPROM 和 FLASH 存储器基本上相同的处理步骤形成的逻辑电路。
3. 权利要求 2 的晶片，其中所述逻辑电路包括处理器。
4. 权利要求 3 的晶片，其中所述逻辑电路包括总线接口和本地
10 总线，所述存储器被耦合到所述本地总线。
5. 权利要求 1 的晶片，其中至少一个所述存储器包括浮栅，以及通过使用衬底热电子注入，把电荷注入到所述浮栅。
6. 权利要求 1 的晶片，其中两种所述存储器都使用衬底热电子
注入。
7. 权利要求 1 的晶片，其中所述 EEPROM 存储器实行字节擦除。
8. 权利要求 1 的晶片，其中所述 FLASH 存储器实行块擦除。
9. 权利要求 1 的晶片，其中所述 EEPROM 和 FLASH 存储器包括单
元，所述 FLASH 存储器的所述单元小于所述 EEPROM 存储器的所述单元。
10. 权利要求 1 的晶片，其中所述单元包括双层多晶硅传感晶体
20 管。
11. 权利要求 1 的晶片，其中所述 FLASH 和 EEPROM 单元包括选择
晶体管。
12. 形成集成电路晶片的方法，包括：
在所述晶片上形成 EEPROM 存储器；
在所述晶片上形成 FLASH 存储器；以及
25 使用基本上相同的处理步骤形成所述 EEPROM 和 FLASH 存储器。
13. 权利要求 12 的方法，包括使用完全相同的处理步骤形成所述
EEPROM 和 FLASH 存储器。
14. 权利要求 12 的方法，包括使用衬底热电子注入以便把电荷注
30 入到所述存储器的浮栅。
15. 权利要求 12 的方法，包括对于所述 EEPROM 存储器使用字节
擦除。



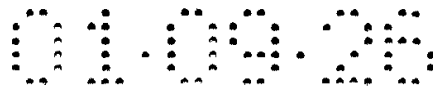
16. 权利要求 12 的方法，包括对于所述 FLASH 存储器使用块擦除。

17. 权利要求 12 的方法，包括形成所述 EEPROM 和 FLASH 存储器，以使得所述 FLASH 存储器的所述单元小于所述 EEPROM 存储器的所述单元。

5 18. 权利要求 12 的方法，还包括通过使用与形成所述 EEPROM 和 FLASH 存储器的基本上相同的处理步骤，以便在所述晶片上形成逻辑电路

19. 权利要求 18 的方法，包括使用完全相同的处理步骤来形成所述逻辑电路和所述存储器。

10 20. 权利要求 18 的方法，其中形成所述逻辑电路包括形成处理器。



说明书

带有嵌入的快闪和 EEPROM 存储器的器件

背景

5 本发明总的涉及包括嵌入的非易失性存储器的器件。

非易失性存储器单元由于它们即使在存储器的电源关断时仍旧能保持记录的信息因而十分有好处。有几种不同类型的非易失性存储器，包括可擦除可编程只读存储器（EPROM）、电可擦除可编程只读存储器（EEPROM）、和快闪（flash）EEPROM 存储器。EPROM 是可通过曝光而擦除的，但可通过沟道热电子注入到浮栅上而被电编程。传统的 EEPROM 具有相同的编程功能，但代替光可擦除作用，它们可以通过电子隧道效应被擦除和被编程。因此，信息可被存储在这些存储器中，当电源关断时信息被保持，以及如果必要的话，使用适当的技术时存储器就可被擦除以便重新编程。快闪 EEPROM 可被成块地被擦除，典型地比常规
15 的 EEPROM 给出较佳的读存取时间。

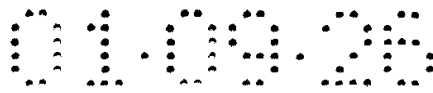
当前，快闪存储器的使用很流行。例如，快闪存储器常常被利用于希望在其中存储需要快速更新的代码的微控制器、调制解调器和智能（SMART）卡等之中提供芯片上存储器。

20 虽然快闪存储器和 EEPROM 是紧密有关的，在许多场合下，快闪存储器是优选的，因为它们的更小的单元尺寸意味着它们能做得更经济。然而，快闪存储器和 EEPROM 常常具有非常类似的单元属性。

非易失性存储器单元在某些方面不同于在所谓的逻辑装置的电子部件（诸如连同存储器单元一起工作的微控制器）中所利用的晶体管。逻辑装置由使用单个栅极的晶体管形成。非易失性存储器通常包括两个栅极，被称为控制栅和浮栅，其一个放置在另一个上。因为这种结构上的差别，非易失性存储器和逻辑装置可以通过不同的处理过程制成。这会在处理复杂性和制造成本上造成很大的增加。

30 特别是对于 EEPROM，各个单元进行电编程通常需要向这些单元加上很显著的电位。这些电位包括从 N+ 区域到浮栅的电子隧道效应。比起通常的晶体管运行所需要的电压来说，如果需要提供更高的电压给存储器单元，就会造成附加的复杂性。

虽然在工业上已趋向于对于逻辑电路和非易失性存储器的处理技



术要加以区分开，以及虽然在工业上人们已认识到对快闪 EEPROM 编程需要很大的电流，但对于不需要专门的处理技术或不需要相对较高的编程电压和较高的电流的电可擦和可编程非易失性存储器有重大的要求。

5 而且，对于传统的 FLASH EEPROM，对各个单元进行电编程需要向这些单元施加很高的电流。这样，非常小量的这种电子流就从漏极耗尽区注入浮栅。这意味着，这样的器件的注入效率是低的（例如， 1×10^{-9} ）。由于运行在低电压下的高的电流泵的设计，高电流的需求增加了附加的复杂性。

10 传统上，利用三种方法来把 FLASH 和 EEPROM 集成在单个集成电路晶片上。一种方法是通过使用适当的处理技术建造 EEPROM 和 FLASH 器件，以便把两种不同类型的器件产生在同一个晶片上。然而，这导致所涉及的处理步骤数目的很大的增加，所以大大地增加了所得到的器件的成本。所以，这样的技术在工业界不能得到显著地接受。

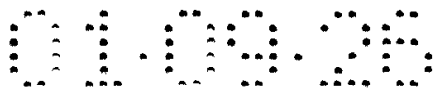
15 替换地，可以产生一种基本的 FLASH 存储器，以及使一个附加的 FLASH 部分设计成适合于模仿 EEPROM 存储器。这通常涉及把软件制做在 FLASH 存储器中，以使得 FLASH 能够作为 EEPROM 存储器运行。该软件被存储在一个也是 FLASH 存储器的引导块中。因此，该系统需要一个作为 FLASH 的第一 FLASH 存储器，一个用于存储对于模仿 EEPROM 运行所需要的软件的第二 FLASH 存储器，以及一个能实际实施像 FLASH 20 那样的能力的附加 FLASH 存储器。这导致非常昂贵的结构，它的运行很复杂。因此，这种技术从管理学的角度也不能被接受。

第三种方法是使用 EEPROM 存储器来模仿 FLASH 存储器。然而，EEPROM 存储器通常很大，所以显得昂贵得多。事实上，EEPROM 存储器 25 比 FLASH 存储器大三到四倍。所以，这种方法通常认为商业上是不可行的，同样得不到市场的接受。

因此，需要继续寻求把 EEPROM 和 FLASH 能力建造在单个集成电路晶片上的方法。

发明概要

30 按照一个实施例，集成电路晶片包括第一部分，其上包括逻辑电路；第二部分，其上包括 EEPROM 存储器；以及第三部分，其上包括 FLASH 存储器。



附图简述

图 1 是用于一个实施例的阵列结构的示意图；

图 2 是显示图 1 所示的实施例的一个单元的半导体实施方案的配置的显著地放大的顶视图；

5 图 3 是总的沿图 2 的线 3-3 截取的截面图；

图 4 是本发明的一个实施例的放大顶视图；

图 5 是本发明的一个实施例的方框图；

图 6 是用于一个实施例的阵列结构的示意图；

图 7 是总的沿图 8 的线 7-7 截取的截面图；

10 图 8 是显示图 6 所示的实施例的一个单元的半导体实施方案的配置的很大地放大的、顶视图；

图 9 是图 8 所示的单元的部分三维视图；

图 10 是另一个实施例的截面图。

优选实施例描述

15 逻辑器件 100 可以被制造在一个其中也包括嵌入的 FLASH 和 EEPROM 存储器 102 和 104 的晶片上。这样，以同一种处理技术制做的逻辑器件和两种存储器器件，可以以紧凑的形式被集成在一起。在某些实施例中，这种形式具有成本和速度上的优点。

20 例如，EEPROM 存储器 104 在单元尺寸上可能较大，以及可以提供字节擦除。FLASH 存储器在单元尺寸上可能较小，以及可以进行块擦除。例如在一个例子中，逻辑电路可以是用于蜂窝电话的逻辑电路，其中包括：处理器 106，如图 5 所示，它使用 EEPROM 存储器 104，用来存储经常改变的电话号码；以及 FLASH 存储器 102，用来存储很少改变的协议。

25 在一个实施例中，如图 5 所示，存储器 102 和 104 可以通过本地总线 112 和接口 110 被耦合到处理器 106 和系统存储器控制器 108。控制器 108 可以控制芯片以外的随机接入存储器（以虚线表示）。

30 如图 1 所示，EEPROM 存储器单元 10 包括传感晶体管 12 和选择晶体管 14。这个结构被有利地实施在半导体层上，在该层上设置有绝缘的浮栅 22。

对于每个单元 10a-10d，选择晶体管 14 的源极 13 被源极节点 56 控制。选择晶体管 11 的栅极被节点 51 控制。传感晶体管 12 的控制栅

27 被控制节点 57 控制。传感晶体管 12 的漏极 16 被连接到漏极节点 55。

如图 2 所示，用于实施单元 10 的一个配置包括控制栅 27。控制栅 27 延伸跨过有源区 18，它以传感晶体管 12 的漏极 16 和选择晶体管 14 的源极 13 为边界。选择栅 11 也平行地延伸，以及穿过控制栅 27，到达控制栅 27 的边缘和区域 15a。控制栅 27 可以非自对准选择栅 11 和传感栅极 12。浮栅 22 也被设置在有源区 18 之上以及在控制栅 27 的下面被隔离开。

漏极 16 可以包括触点 55，如图 2 所示，它被连接到漏极扩散区 16。源极节点 56 也可以由一个触点来实现。

图 3 上显示了传感晶体管 12 与选择晶体管 14 之间的关系。浮栅 22 形成了晶体管的一些部分，该晶体管具有漏极 16 和源极 13。同样地，选择栅 11 形成在晶体管的源极 13 与漏极 16 之间的其它部分。传感晶体管 12 包括沟道 25a，而选择晶体管 14 包括沟道 24。控制栅形成电容的平板，它的沟道是 15a。选择栅 11、浮栅 22 和控制栅 27 形成带有源极 13 和漏极 16 的晶体管的栅极。

在显示的实施例中，沟道 25a 和 24 是 P 型半导体材料，以及是 P 阱 28 的一部分。P 阱 28 又被形成在 N 阱 29 中。最后，N 阱 29 被形成在 P 型衬底 38 中。P 阱 28 可以被加上一个偏压，如在 70 处所示的，以及 N 阱 29 可以被加上偏压，如在 72 处所示的。

在控制栅 27 重叠在选择栅 11 和浮栅 22 之间的衬底区域 15a 的地方形成了电容 50，它控制区域 15a 的耗尽/反型区，造成由传感晶体管 12 形成的耗尽/反型区域 25 的扩展。这是在编程和读出操作期间运行的。在读操作中，电容 50 通过形成一个反型区从而桥接传感和选择晶体管沟道。当反型区在区域 15a 下被形成时，传感和选择晶体管 12 和 14 被连接。

浮栅 22 通过它与沟道 25a 的交互作用而形成隧道电容 33。隧道氧化层 30 把浮栅 22 与沟道 25a 分开。同样地，共聚介质氧化层 40（它是耦合电容 32 的一部分）把浮栅 22 与控制栅 27 分开。最后，由氧化层 51 把控制栅 27 与区域 15a 分开。同样地，选择晶体管 14 包括栅极氧化层 52，它可以具有与隧道氧化层 30 相同的厚度。

控制栅 27 与选择栅 11 的重叠是为了处理的便利。同样地，控制栅 27 被显示为与漏极 16 重叠，但这也仅仅是为了处理的便利。控制

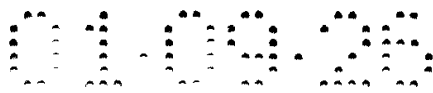
栅 27 不需要自对准传感或选择晶体管。

单元 10 可被描述为一个利用用于编程的高效率的衬底热电子注入和用于擦除的 Fowler-Nordheim 隧道效应的 EEPROM。衬底热电子注入的处理过程在以下的论文中有很好的描述：T.H.Ning, C.M, Osburn 和 J.W. Yu, “Emission Probability of Hot Electrons from Silicon into Silicon Dioxide (热电子从硅到二氧化硅的发射概率)”, J. Appl. Phus., vol.48, p.286, (1977); Boaz Eitan, James L. McCreary, Daniel Amrany, Joseph Shappir, “Substrate Hot-electron Injection EPROM (衬底热电子注入 EPROM)”, IEEE Transactions on Electron Devices, Vol. ED-31, No. 1, p. 934, (July 1984); I. C. Chen, C. Kaya, and J. Paterson, “Band-to-Band Tunneling Induced Substrate Hot-electron (BBISHE) Injection: A New Programming Mechanism for Nonvolatile Memory Devices (带到带隧道效应感应的衬底热电子注入: 用于非易失性存储器器件的新的编程机制)”, IEDM (1989) p. 263; 和 C. Y. Hu, D. L. Kencke, S. K. Benerjee, “Substrate-current-induced Hot-electron (SCIHE) Injection: A New Convergence Scheme for FLASH Memory (衬底电流感应的热电子注入: 用于 FLASH 存储器的新的会聚方案)”, IEDM (1995) p. 283. 这些论文中的每一篇都直接在此引用, 以供参考。

编程是通过高效率的衬底热电子注入而达到的。如图 3 所示, 衬底电子 (用在 60 处的箭头表示) 是通过给源极 13 加正向偏置而产生的, 源极 13 是由选择晶体管沟道 24 和在电容 50 下面的区域 15a 而与传感晶体管 12 沟道 25a 分开的。某些衬底电极 60 扩散通过沟道 24 下面的区域, 到传感晶体管 12 下面的区域 25a。

对于需要被编程的单元, 沟道区域 25a 被加偏置, 以使得形成耗尽区 25。耗尽区 25 通过电容 50 的运行, 也在电容 50 之下被扩展, 如在 15 处表示的。当电子达到耗尽区 25 时, 它被电场 V_{cs} 加速。电场 V_{cs} 是沟道 25a 的电位 (表面反型区的电位) 与 P 阱 28 的电位之间的差值。这些电子中某些电子得到了超过有效的氧化层势垒高度电位的足够能量, 从而被注入到浮栅 22。

对于不要被编程的单元, 沟道到 P 阱的电位小于有效的氧化层势垒高度。在这种情形下, 电子没有得到足以克服势垒高度的能量, 以



及不被注入到浮栅 22。

N+掺杂区 13、在选择晶体管 14 与传感晶体管沟道 25a 之下的 P 区 24、以及在电容 50 之下的耗尽区 15 形成了横向双极性晶体管 62。双极性晶体管的发射极（源极 13）用作为电荷注入器，把衬底电子从源极扩散区注入到处在浮栅 22 下的偏置耗尽区。通过以扩散区 13 作为发射极和沟道 24 作为基极，收集极是偏置耗尽区 25（包括区 15）。由于沟道区 25a 在读期间用作为对于传感晶体管的沟道，以及在传感晶体管 12 之下的偏置耗尽区 25 在编程期间用作为双极性晶体管 62 的收集极，所以得到紧凑的单元配置。

衬底热电子注入的效率是多个特性的函数。考虑耗尽区 25，电子以晶格声子散射的方式通过具有一定的电子平均自由路径的耗尽区 25 进行散射。这些电子中的某些电子由于没有被较多地散射，所获得的能量足以克服有效的势垒高度，以及被注入到浮栅 22。某些电子得到比有效的势垒高度较少的能量，以及不被注入到浮栅 22。注入效率是一个与掺杂特性和沟道到 P 阱的电位 V_{cs} 的相关性很强的函数。

由于单元 10 位于被嵌入在 N 阱 29 中的 P 阱 28 中，在编程期间，通过把控制栅 27 上升到 V_{pp} （它可以是 7 到 14 伏），浮栅 22 经过耦合电容 32 被容性耦合到更高的电压。浮栅 22 达到一个低的漏极偏置情况下的电压近似地为当控制栅 27 和 P 阱 28 与漏极 16 处在地电位时的浮栅电压加上耦合比乘以控制栅 27 上的电压的函数。所述耦合比在一阶近似时等于电容 32 的电容值除以耦合电容 32 和隧道电容 33 的电容值的总和。

当选择晶体管 14 关断时，传感晶体管漏极 16 的电位可被强制地接近于电源电位 V_{cc} 或更高。由于选择晶体管 14 被关断，节点 51 的电位跟随沟道 25a 的电位。沟道 25a 的电位就是沟道区 25a 的表面反型区的电位，它被按照如下方式来设置。当浮栅 22 的电位 (V_{fg}) 比漏极 16 的电位高出一个传感晶体管 12 的门限电压时，沟道电位与漏极电位相同。另一方面，当浮栅 22 的电位小于漏极 16 的电位加上传感晶体管 12 的门限电压时，沟道电位是浮栅 22 的电压与传感晶体管 12 的门限电压之间的差值。

P 阱的电位是加到 P 阱 28 的电压 70。由于 P 阱 28 被嵌入在 N 阱 29 中，以及 N 阱被设置为近似于 V_{ss} 或更高的电压 72，P 阱的电位 V_p 可

以是负的，典型地是-1 到-2 伏。而且，它通常小于有效的氧化层势垒高度，以便避免任何电位扰动问题。

在沟道 25a 区与 P 阱 28 的电位 (V_p) 70 之间的电位差是跨在耗尽区 25 上的电压。对于要被编程的单元，漏极 16 的电压被升高，典型地接近 V_{cc} 或更高。在传感晶体管 12 和电容 50 之下的沟道 25a 和 24 中的耗尽区 25 可以用一个等于沟道电位减去 P 阱的电位 70 的电压降来形成。

对于不要被编程的单元，漏极 16 的电压被设置为 0 伏 (V_{ss})。在耗尽区 25 的电压降等于 V_p 的绝对值，它典型地小于有效的氧化层势垒高度。

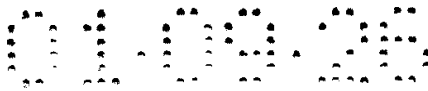
单元 10 的擦除是通过电子从浮栅 22 到沟道区 25a 和漏极扩散区 16 的 Fowler-Nordheim 隧道效应而达到的。在擦除期间，控制栅 27 被强制为从-7 到-14 伏的负电压。至于漏极扩散区 16、P 阱 28、和 N 阱 29，它们被加偏压到接近于 V_{cc} 或更高的正电位。 V_{cc} 是由所利用的特定的技术确定的。例如，利用本技术，它可以是 5.0 到 2.5 伏。这减小了在 N+ 扩散区 16 与 P 阱 28 之间结上的电场。减小了的电场阻止在浮栅 22 下面的栅极氧化层中热空穴的加速俘获。

优选地不把漏极 16 的偏压加到高于 P 阱 28 的电压从而达到这样的程度，以致于使得栅极感应的漏极泄漏 (GIDL) 成为一个问题。对于当前的技术，这意味着漏极 16 的偏压不可能比 P 阱 28 的偏压高出约 1 伏。另外，如果漏极 16 的偏压大大地超过 P 阱 28 的偏压，则由于横向结场的加速度，在选择栅极氧化层 52 中会发生热空穴俘获。

因为 P 阱 28 被嵌入在 N 阱 29 中，这使得把正电压加到 P 阱的能力上升。P 阱的电压优选地等于或小于 N 阱的电位，以避免 P 阱/N 阱正向偏置。因此，把 V_{cc} 或更高的正电压加到 P 阱、N 阱和漏极 16，可以消除由 GIDL 引起的热空穴俘获，而同时允许漏极 16 的电压上升到 V_{cc} 或更高。

在电容 33 上的电压是在一方面的浮栅 22 的电位与扩散区 16 和 P 阱 28 的电位之间的差值。当电位超过 8 到 10 伏时，产生足够的隧道电流，以及根据隧道氧化层 30 的厚度，浮栅 22 可以在几毫秒到几秒的时间范围内被擦除到负的电位。

电子从隧道到达漏极区 16 (漏极擦除)。隧道电流取决于从浮栅 22



到漏极 16 的电压。在本发明的一个实施例中，使用字节擦除。

读单元 10 的编程状态可以如下地完成。通过强制控制栅 27 成为正的电位（例如 2.5 到 5 伏），浮栅 22 被容性耦合到更高的电位。浮栅 22 被耦合到电位 V_{fg} ，它可被计算为等于当控制栅 27 保持为接地时的浮栅电位加上在控制栅 27 上的电位乘以耦合比之和。

在读操作期间的漏极 16 的电位被限制为小于 2 伏的电压。这可避免任何读干扰。

对于要被读出的被选择的单元，选择栅 11 被强制成为 V_{cc} ，以及源极 13 被强制成为接地。未被选择的选择栅 11 也被强制成为接地。

当这些电位被加到选择的单元 10 时，电流流过传感晶体管 12。这个电流然后被馈送到一个电流传感放大器（未示出）。如果在浮栅 22 上的电压大于在传感晶体管 12 上的门限电压，则一个较高的电流（或许大于 20 微安）被检测为导通状态。当浮栅的电位小于门限电压时，则将会流过较低的电流（例如小于 20 微安），以及检测到非导通状态。

被检测的导通状态可被称为“1”状态。非导通状态可被称为“0”状态。

一个关于在编程、读出和擦除时单元运行的例子可以被概括在以下的图表中：

单元运行

	擦除		编程		读数	
	选择	未选择	选择	未选择	选择	未选择
漏极 (55)	$\geq V_{cc}$	*0 或 $\geq V_{cc}$	$\geq V_{cc}$	*0 或 $\geq V_{cc}$	-1.5V	*0 或 -1.5V
源极 (56)	浮动	浮动	V_s	浮动	V_{ss}	V_{ss}
选择 (51)	V_{ss}	V_{ss}	V_{ss} 到 V_s	V_{ss} 到 V_s	V_{cc}	V_{ss}
控制 (57)	-7 到 -14V	V_{ss}	V_{pp}	V_{ss}	2 到 5V	V_{ss}
N-阱 (29)	$\geq V_{cc}$	$\geq V_{cc}$	V_{cc} 到 V_{ss}	V_{cc} 到 V_{ss}	V_{cc}	V_{cc}
P-阱 (29)	$\geq V_{cc}$	$\geq V_{cc}$	V_{bias}	V_{bias}	V_{ss}	V_{ss}

*0 是对于未选择的列。

$V_{pp} \approx 7$ 到 14 伏。

V_s 是由注入电流量设置的节点电压，其电流范围从几十纳安到几十微安，这取决于编程速度要求。典型地，它是从几十毫秒到几十微秒。 V_{bias} 是 P 阱 28 上的偏压，它可以是 V_{ss} 或它可被强制成为 -1 到 -2 伏，以便提高注入效率。

- 5 用于产生两个负的偏置电位（一个给控制栅 57 加偏压，另一个给 P 阱 28 加负的偏压）的适当的芯片上电路可在以下的著作中找到：
L. A. Glasser 和 D. W. Dobberpuhl, "The Design and Analysis of VLSI Circuits (超大规模集成电路的设计和分析)", (Dec. 1985), Addison-Wesley 出版, pp. 301-329, 该著作直接在此引用, 以供参考。
10 V_{ss} 是外部地电位。

虽然单元 10 可以作为单个元件被利用，但它也可被连接为阵列，如图 1 所示。在阵列中，显示了多个单元 10, 10a, 10b, 10c, 10d。源极节点 13 是通过连接在同一个行中所有的单元的源极节点作为一个单个节点 56 而被形成的。控制节点 17 是通过连接在同一个行中各个
15 单元的所有的控制节点在一起作为一个单个节点 57 而被形成的。选择栅节点 11 是通过连接在同一个行中所有的单元的选择栅节点在一起作为一个节点 51 而被形成的。同样地，漏极节点 16 是通过连接在同一列中所有单元的漏极节点在一起作为一个单个节点 55 而被形成的。这个节点被引到传感放大器（未示出）。

20 用于 FLASH 存储器 102 的一对 FLASH 存储器单元 210a 和 210b（如图 6 所示）位于两个相邻的字线 212a 和 212b 中。两个单元 210 被安排在同一比特线 214 上。其它的单元 216, 218, 和 220 被安排在比特线 222, 224 和 226 元件字线 212 上。这种结构在一个其上设置了电绝缘的浮栅 228 的半导体层上有利地被实施。

25 每个单元 210, 216, 218 和 220 的源极端子 230 被源 132 或 234 控制。单元 210, 216, 218 和 220 的控制栅 236a 和 236b 被字线 212 控制。单元的漏极 238 被连接到比特线 214, 222, 224 和 226。

图 7 所示的、用于实施一对单元 210 的一种配置包括一对控制栅 236a 和 236b。同一个配置可被使用于其它的单元对 216, 218 和 220。
30 控制栅 36 扩展到场氧化层隔离区 240, 后者在一边以漏极 238 为边界以及在另一边以源极 232 和 234 为边界。浮栅 228a 和 228b 被设置在氧化层隔离区 240 上与控制栅 236a 和 238a 相隔离, 并处在控制栅 236a

和 238a 之下。

在显示的实施例中，在比特线下的有源区是 P 型半导体材料以及是 P 阱 244 的一部分。P 阱 244 又被形成在 N 阱 246 中。最后，N 阱 246 被形成在 P 型衬底 248 上。P 阱 244 可以通过加重掺杂的 P 型触点 252 而被加偏压（如在 250 处所表示的），以及 N 阱 246 可被加偏压（如在 254 处所表示的）。

每个浮栅 228 通过它与沟道 258 交互作用而形成隧道电容 257。隧道介质 256 把浮栅 228 与沟道 258 分开。同样地，作为耦合电容 261 的一部分的共聚介质 260 把浮栅 228 与控制栅 236 分开。最后，由介质 262 把控制栅 236 与沟道 264 分开。

图 8 上显示单元 210, 216, 218 和 220 的物理关系。浮栅 228 以虚线显示。同样地，控制栅 236 跨在漏极 238 与源极 232 或 234 之间。漏极触点 242 位于控制栅 236 与比特线 214, 222 和 226 之间。比特线 214, 222, 224 和 226 横向延伸到控制栅 236。

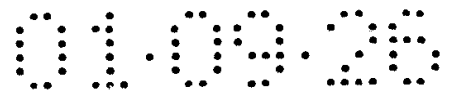
单元 210 可被描述为一个利用用于擦除的高效率的衬底热电子注入和用于编程的 Fowler-Nordheim 隧道效应的快闪 EEPROM。

各个单元是通过使得电荷积累在浮栅 228 上而被擦除的，以及可以在电荷从浮栅被去除时达到编程。因此，一开始所有的单元在它们的浮栅上累积了电荷。所有的单元 210, 216, 218 和 220 通过使用块擦除方式从而一起被擦除。

擦除是通过高效率衬底热载流子注入而实现的。如图 7 所示，衬底载流子（诸如电子，在 266 处以箭头表示）可以通过给源极 232 或 234 加正向偏压而被产生。也就是，一个单元的源极被用作为用于一对单元 210, 216, 218 或 220 中的另一个单元的注入器。因此，源极 232 用作为图 7 上被擦除的单元 210b 的注入器。同样地，当单元 210a 被擦除时，源极 234 被用作为用于单元 210a 的注入器。

由沟道 264a 和 258 以及漏极 238 把源极 232 与单元 210b 分开。某些衬底电子 266 通过这些区扩散到单元 210b 下面的沟道区 256b。漏极 238 低效率地取得一些电子。

对于需要被擦除的单元，沟道区 256 可以被加偏压，从而形成耗尽区。当电子达到耗尽区时，它被电场 V_{cs} 加速。电场 V_{cs} 是沟道电位（表面反型区的电位）与 P 阱 244 的电位之间的差值。这些电子中



的某些电子得到超过有效的氧化层势垒高度电位的足够能量，从而被注入到浮栅 228b。

5 对于不需要被擦除的单元，沟道到 P 阱的电位小于有效的氧化层势垒高度。在这种情形下，电子没有得到足以克服势垒高度的能量，以及不被注入到浮栅 228。

在希望快闪擦除的情形下，存在这样一种方案，其中在同一个 P 阱内的所有的单元可以一起被擦除。这是通过给 P 阱 244 相对于 N 阱 246 加正向偏压而达到。电子从 P 阱 244 的底部被注入（如箭头 280 所表示的），以及被浮栅或漏极收集（如图 10 所示）。

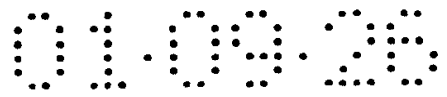
10 加重掺杂的 N 型源 232、在单元 210a 之下的 P 型区 264a 和 256a、以及在漏极 238 与单元 210b 沟道 256a 之下的区形成横向双极性晶体管 268。双极性晶体管 268 的发射极（源极 232）被用作为一个电荷注入器，把衬底电子从源极注入到在浮栅 228a 之下的偏置耗尽区。通过将源极 232 作为发射极和沟道 268a, 256a, 以及将在漏极 238 下的区
15 作为基极，则收集极是偏置耗尽区 256b。

因为分开的选择晶体管是不必要的以及漏极 238 被用作为用于一对单元的两个相邻的单元的漏极，所以得到紧凑的单元配置。一个单元的源极也被用作为用于一对单元中的另一个单元的有效注入器。

20 衬底热电子注入的效率是多个特性的函数。考虑耗尽区 256b，电子以晶格声子散射的方式通过具有一定的电子平均自由路径的耗尽区 25 上散射。这些电子中的某些电子由于没有被较多地散射，所获得的能量足以克服有效的势垒高度，以及被注入到浮栅 228。某些电子得到比有效的势垒高度较少的能量，以及不被注入到浮栅 228。注入效率是一个与掺杂浓度和沟道到 P 阱的电位 V_{cs} 的相关性很强的函数。

25 由于单元 210 位于被嵌入处在 N 阱 246 中的 P 阱 244 之中，在擦除期间，通过提升比特线上的电位（它可被加偏压到从 7 到 14 伏的电位），浮栅 228 被容性耦合到高电压。浮栅 228 达到一个低的漏极偏置情况下的电压近似地为当控制栅 236b 和 P 阱 244 与漏极 238 处在地电位时的浮栅上的电压加上耦合比乘以控制栅 236b 上的电压的函数。该
30 耦合比在一阶近似时等于耦合电容 261 的电容值除以耦合电容 261 和隧道效应电容 257 的电容值的总和。

当单元 210a 关断时，漏极 238 电位可被强制地接近于电源电位 V_{cc}



或更高。沟道 256a 的电位是沟道区的表面反型区的电位，它被设置为如下所述的那样。当浮栅 228 的电位 (V_{fg}) 比漏极 238 的电位高出一个单元门限电压时，沟道电位与漏极电位相同。另一方面，当浮栅 228 的电位小于漏极 238 的电位加上单元门限电压时，沟道电位是浮栅 228 的电压与单元门限电压之间的差值。

P 阱的电位是加到 P 阱 244 的电压 250。P 阱 244 被嵌入在 N 阱 246 中，以及 N 阱被设置为近似于 V_{ss} 或更高的电压 254，诸如 V_{cc} 。因此，P 阱的电位 V_p 可以是 V_{ss} 或负的，典型地是 -1 到 -2 伏。而且，P 阱的电位通常小于有效的氧化层势垒高度，以避免任何电位扰动问题。

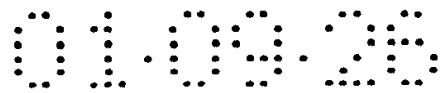
在沟道 256b 区与 P 阱 244 的电位 250 之间的电位差是跨在耗尽区上的电压。对于要被擦除的单元，漏极 238 的电压被升高，典型地接近 V_{cc} 或更高。在要被擦除的单元下面的耗尽区可以用一个等于沟道电位减去 P 阱的电位 250 的电压降来形成。

对于不要被擦除的单元，未选择的字线电压可被设置为 0 伏 (V_{ss})。耗尽区上的电压降小于有效的氧化层势垒高度。

单元 210 的编程是通过电子从浮栅 228 到沟道区 256b 和漏极 238 的 Fowler-Nordheim 隧道效应而达到的。在编程期间，选择的比特线 236b 被强制成为约 5 伏的高电压“H”（高于 V_{cc} ）。未选择的比特线被保持在 V_{ss} 。如果 N 阱和 P 阱分别保持在 V_{cc} 和 V_{ss} ，则漏极 238 和 P 阱 244 之间的结上的电场将被减小。减小的电场阻止在浮栅 22 下的栅极氧化层中热空穴的加速俘获。

电子从隧道到达漏极区 238（漏极编程）。隧道电流取决于从浮栅 228 到漏极 238 的电压。读出单元的编程状态可以如下地完成。被选择的比特线被加偏压大约 0.7 到 1 伏。其余的比特线被允许浮动。源极 232 可被加偏压到 $-V_s$ ，以及源极 234 可被加偏压到 V_H 。被选择的字线被加偏压到 V_{HH} ，以及相邻的字线被加偏压到 $-V_s$ 。P 阱被加偏压到 V_{ss} ，以及 N 阱被加偏压到 V_{cc} 。 V_s 是由注入电流量来设定的节点电压，电流范围从几十纳安到几十微安，这取决于编程速度要求。典型地，它是从几十毫秒到几十微秒。 V_{ss} 是外部地电位。

当这些电位被加到选择的单元 210 时，电流流过该单元。这个电流然后被馈送到电流传感放大器（未示出）。如果在浮栅 228 上的电压大于在传感晶体管 212 上的门限电压，则一个较高的电流（或许大于 20



微安) 被检测为导通状态。当浮栅的电位小于门限电压时, 则流过一个较低的电流 (例如小于 20 微安) 以及检测到非导通状态。

源 232 被加偏压 $-V_s$ (即, 二极管电压降, 约为 0.7 伏), 而源 234 是处在 V_{HH} 。在一个实施例中, V_{HH} 可以约为 10 到 14 伏。选择的字线被加偏压到 V_{HH} , 以及该对的相邻的字线被加偏压到 $-V_s$ 。P 阱的偏压是 $-V_{SS}$, 以及 N 阱的偏压是 V_{CC} 。

在电容 257 上的电压是在一方面浮栅 228 的电位与漏极 238 和 P 阱 244 的电位之间的差值。当差值超过 8 到 10 伏时, 产生足够的隧道电流, 以及浮栅 228 可以在几毫秒到几秒的时间范围内被擦除到负的电位, 这取决于隧道氧化层 56 的厚度。

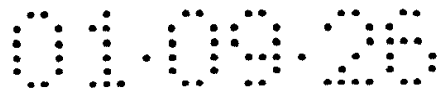
被检测的导通状态可被称为“1”状态。非导通状态可被称为“0”状态。

用于产生两个负的偏置电位 (一个给字线 236 加偏压, 另一个给 P 阱 244 加负的偏压) 的适当的芯片上电路可在以下著作中找到:
L. A. Glasser 和 D. W. Dobberpuhl, “The Design and Analysis of VLSI Circuits (超大规模集成电路的设计和分析)”, (Dec. 1985), Addison-Wesley 出版, pp. 301-329, 该著作直接在此引用, 以供参考。

在阵列中的单元可以通过使用传统的处理过程技术 (诸如双聚合, 单金属 CMOS 处理过程) 而被形成。这里阐述的说明性参量规划一个 0.25 微米或更低的特性尺寸, 采用 1.8 伏的 V_{CC} 电位。当技术允许使用降低的电压和较小的特性尺寸时, 这里的参量随之加以换算。

再次参照图 7, 图上显示电子流从注入器 (它在说明的实施例中是源极 232) 到浮栅 228b 的流动。衬底电子 (用箭头表示) 可以从源极 232 沿着在相邻的隔离区 240 之间的单元的整个宽度 “W” 流动。另外, 电子从扩展到单元宽度以外的部分 232a 和 232b 流动。即, 在宽度方向扩展到单元的任一边的源极 232 的部分也贡献出电子, 它们可被利用来编程浮栅 228。这可大大地增加对于擦除可提供的电子流, 因此可以加快擦除。

这个安排是可能的, 因为加重掺杂的 N 型源 232 和 234 平行于控制栅 236。换句话说, 源 232 和 234 平行于阵列的字线或行。由于存在这样的几何关系, 其长度大于单元的宽度的一部分源极可被利用于擦除的目的。即, 附加的电子可以从单元的相邻列之间的扩展部分 210a



和 210b 流出。

本发明的一个实施例的运行被显示在以下的模式表中：

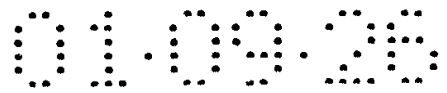
	选择的 B/L	未选择的 B/L	源 1	源 2	仅次于 选择的 WL	选择的 WL
擦除	H	H	-Vs	V _H	-Vs	V _{HH}
编程	H	V _{SS}	浮动	浮动	V _{SS}	-V _{HH}
读出	0.7- 1.0V	浮动	V _{SS}	V _{SS}	V _{SS}	V _{CC}
快闪擦除	H	H	-Vs	-Vs	V _{HH}	V _{HH}

5 阵列中的单元可以通过使用传统的处理过程技术（诸如双聚合，
单金属 CMOS 处理过程）而被形成。这里阐述的说明性参量规划一个 .25
微米或更低的特性尺寸，它采用 1.8 伏的 V_{CC} 电位。当技术允许采用
降低的电压和较小的特性尺寸时，这里的参量随之被加以换算。

10 在本发明的一个有利的实施例中，FLASH 和 EEPROM 存储器可以用
基本上同一个处理步骤被产生。通过减小附加处理步骤的数目，该器
件的总体成本可被大大地改变。在某些实施例中，完全相同的处理步
骤可被利用于两种存储器，这样，不需要额外的其他步骤和额外的掩
膜步骤，在单个晶片上得到两种存储器。

15 开始时使用的衬底材料典型地是 P 型（100）硅，例如，具有范围
为 10-20 欧姆-厘米的电阻率。P 阱 28 和 244 在所谓的三阱处理过程中
被嵌入 N 阱 29 和 246 中。P 阱 28 和 244 具有典型的 4-8 微米的阱的深
度。掺杂浓度可以从 4×10^{15} 到 1×10^{16} 原子/厘米³。三阱由相反掺杂 N
阱 29 和 246 的 P 阱 28 和 244 形成。

20 三阱中单元的形成如下。一个 N 阱植入物通过利用具有 1 到
 1.5×10^{16} 原子/厘米³ 的典型的剂量和 160 到 100Kev 的能量的磷 31 来
构成。N 阱植入物通过使用高温步骤被激发，它典型地是在 1125 到
1150℃ 下经历 6 到 12 小时。N 阱 29 和 246 然后利用 P 阱植入物被反向
掺杂。对于 P 阱植入物的典型的剂量可以是利用诸如硼 11 的物质的 1.5
25 到 2.5×10^{13} 原子/厘米³ 和 30 到 180Kev 的能量。N 阱 29 和 246 以及 P
阱 28 和 244 然后被激发，典型地是在 1125 到 1150℃ 下经历 6 到 12 小



时。这样就可以把阱设置到想要的掺杂浓度和深度。

在阱形成后，应用标准的逻辑场氧化层形成和沟道停止形成步骤。场氧化层厚度和植入物剂量被调节达到 7 到 14 伏的场门限值，它是由用于编程和擦除的 V_{pp} 电平和由逻辑处理能力所确定的。在这个形成
5 后，可以执行存储器单元植入。例如，具有 1.5 到 3×10^{13} 原子/厘米³ 剂量在 30 到 50Kev 的能量下的硼 11 的植入物可以通过牺牲氧化层来实现。例如，85 到 100 埃 (Angstrom) 的干氧化层可以在晶片上生长。干氧化层在部分氧中在 900°C 下生长，随后是 975 到 1050°C 退火。

浮栅 22 和 228 然后可由多晶硅、硅化物或金属形成。如果使用多
10 晶硅，则它的厚度可以是 1600 埃，以及在 870 到 1000°C 下进行 POCl_3 掺杂。共聚介质由氧化物-氮化物-氧化物夹层 (ONO) 形成，下部的氧化层是 60 到 80 埃，氮化物层具有 90 到 180 埃的厚度，以及上部的氧化层是 30 到 40 埃。125 到 200 埃的栅氧化层可以被生长以用于电容 50 之下的氧化层。

15 在规定浮栅和选择栅后，N+植入物被注入到选择晶体管 14 和 210 的源极和传感晶体管 12 和 210 的漏极。它被阻塞在两个栅极之间，这样，它并不进入在电容 50 的极板下面的区域，N+植入物可以是剂量为 1 到 3×10^{13} 原子/厘米³ 及 60Kev 下的磷，后随的是植入物剂量为 2.5 到 4.5×10^{15} 原子/厘米³ 及 60Kev 下的砷。也有可能通过使用轻微掺杂
20 漏极 (LDD) 技术来形成源极和漏极。

如果需要的话，用于控制栅的多晶硅 (poly 2) 然后可被淀积和硅化。栅极通过使用标准的技术被形成图案和被限定。控制栅没有自对准传感和选择栅。

在完成这些电容和晶体管结构的情况下，用于触点和交互连接层
25 的所有随后的处理遵从标准的逻辑后端处理。

虽然在上述的说明中给出多个参量和量值，本领域技术人员将会看到，这些参量和量值仅仅是用于说明的。例如，通过颠倒掺杂结的导电类型和偏压极性，可以实施使用衬底热空穴注入的单元结构。希望由所附属的权利要求覆盖属于本发明的的精神和范围的所有的修改
30 和变动。

说明书附图

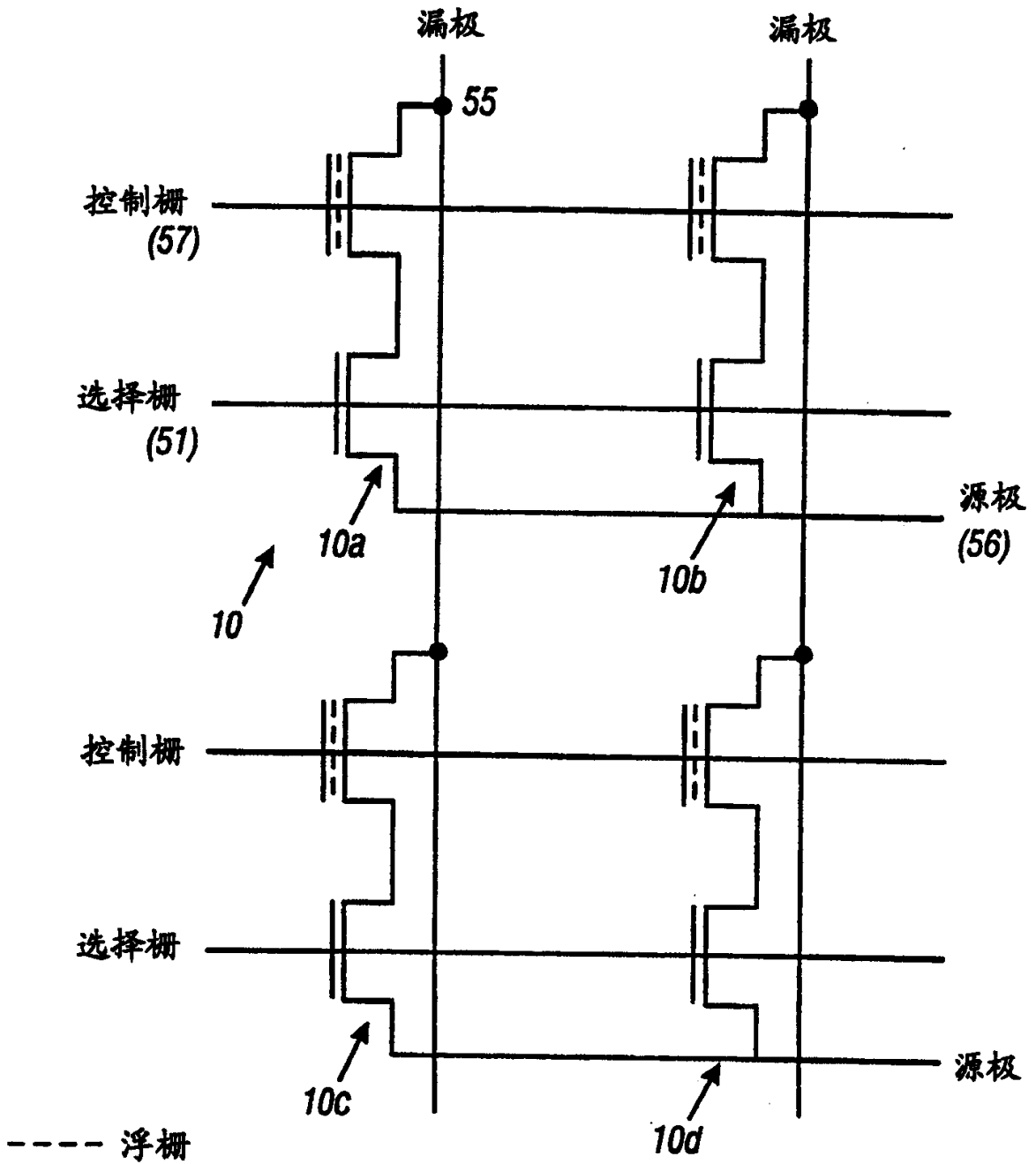


图 1

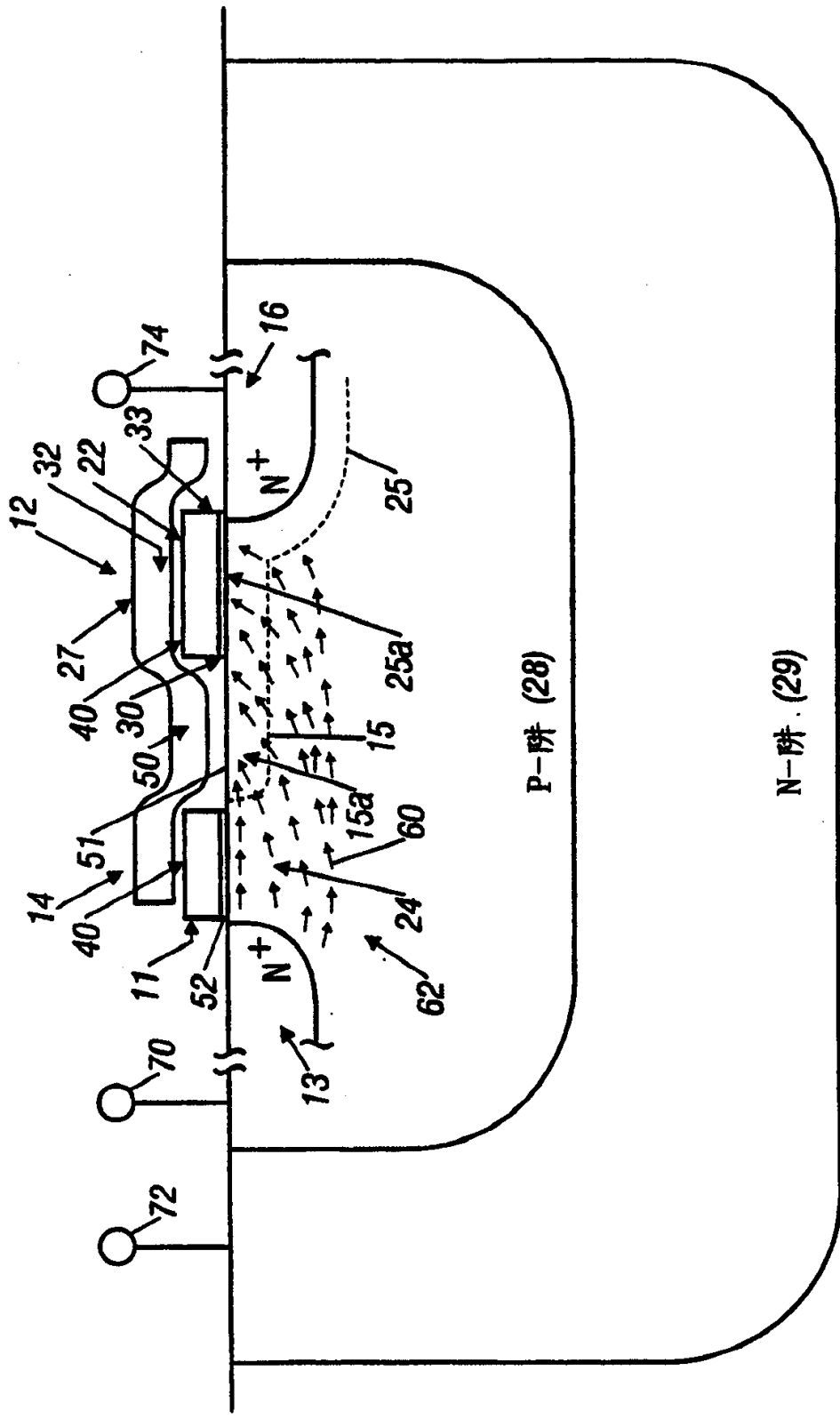


图 3

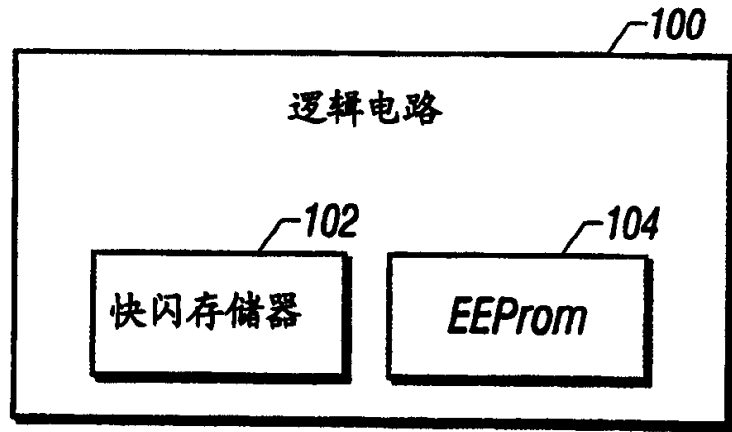


图 4

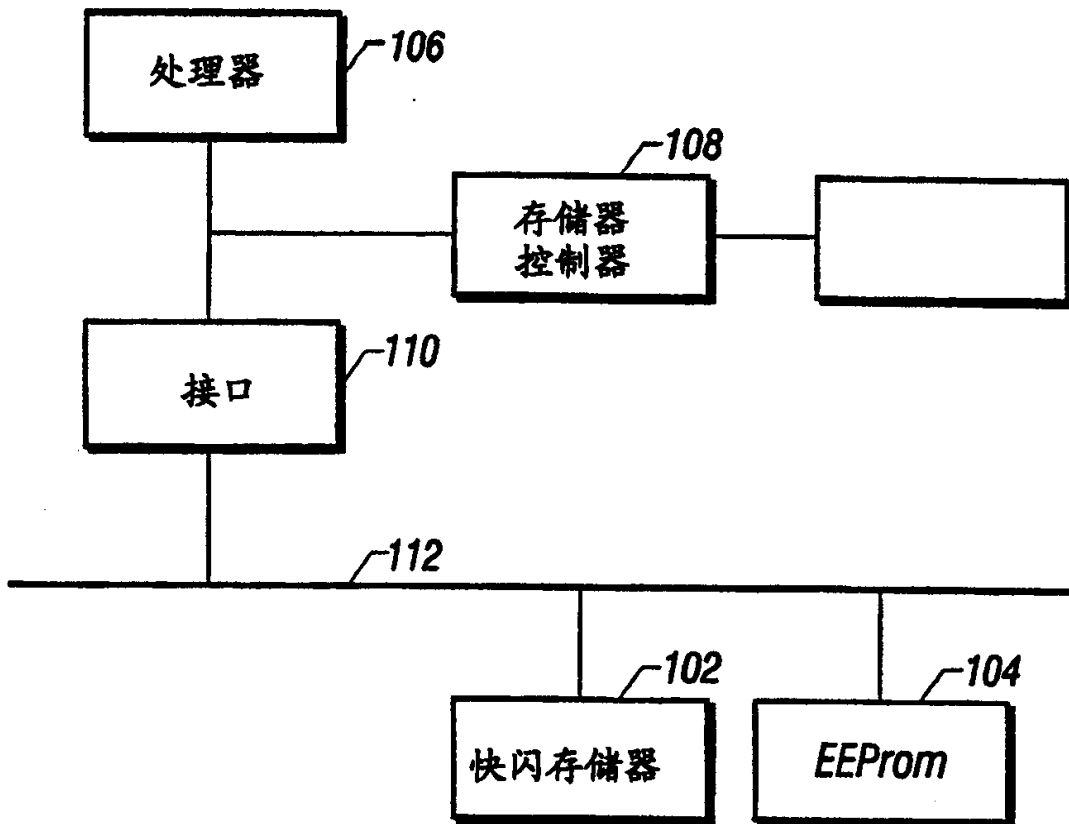


图 5

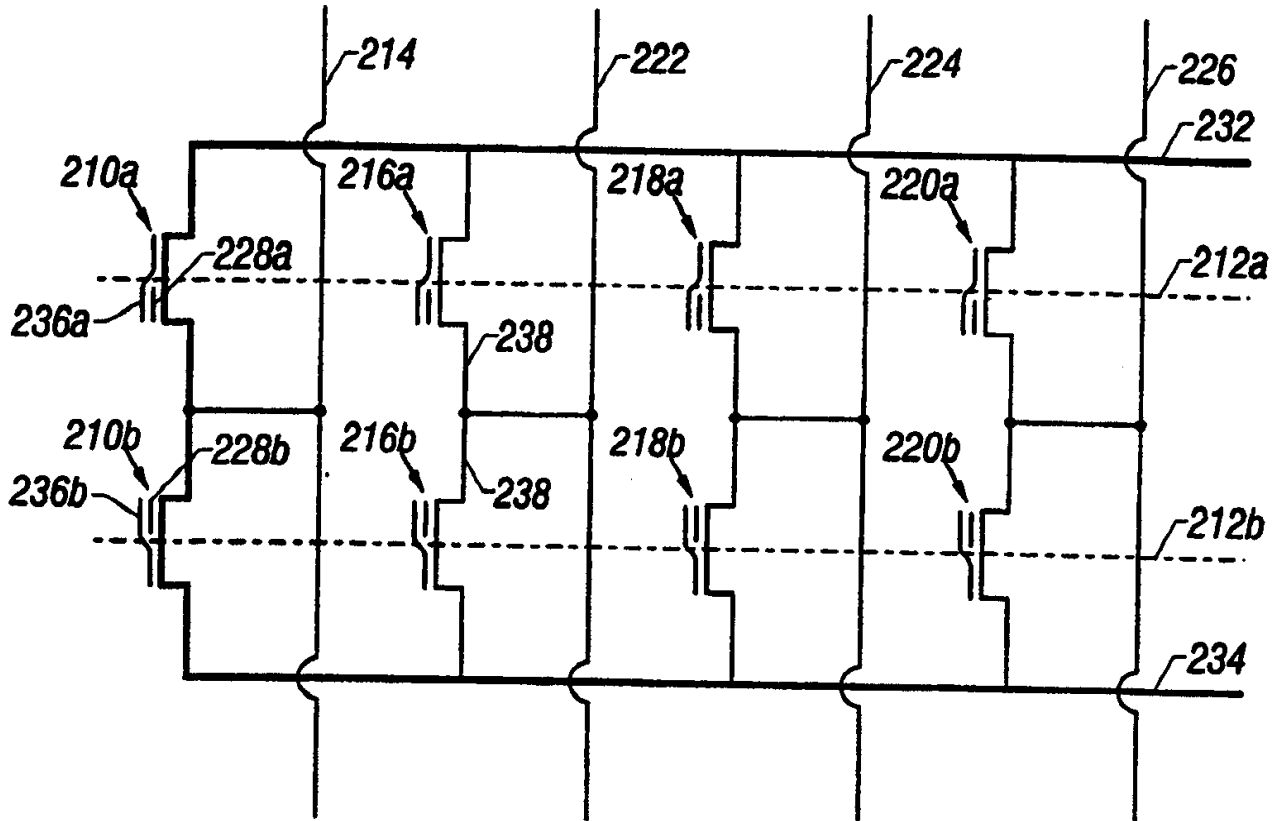


图 6

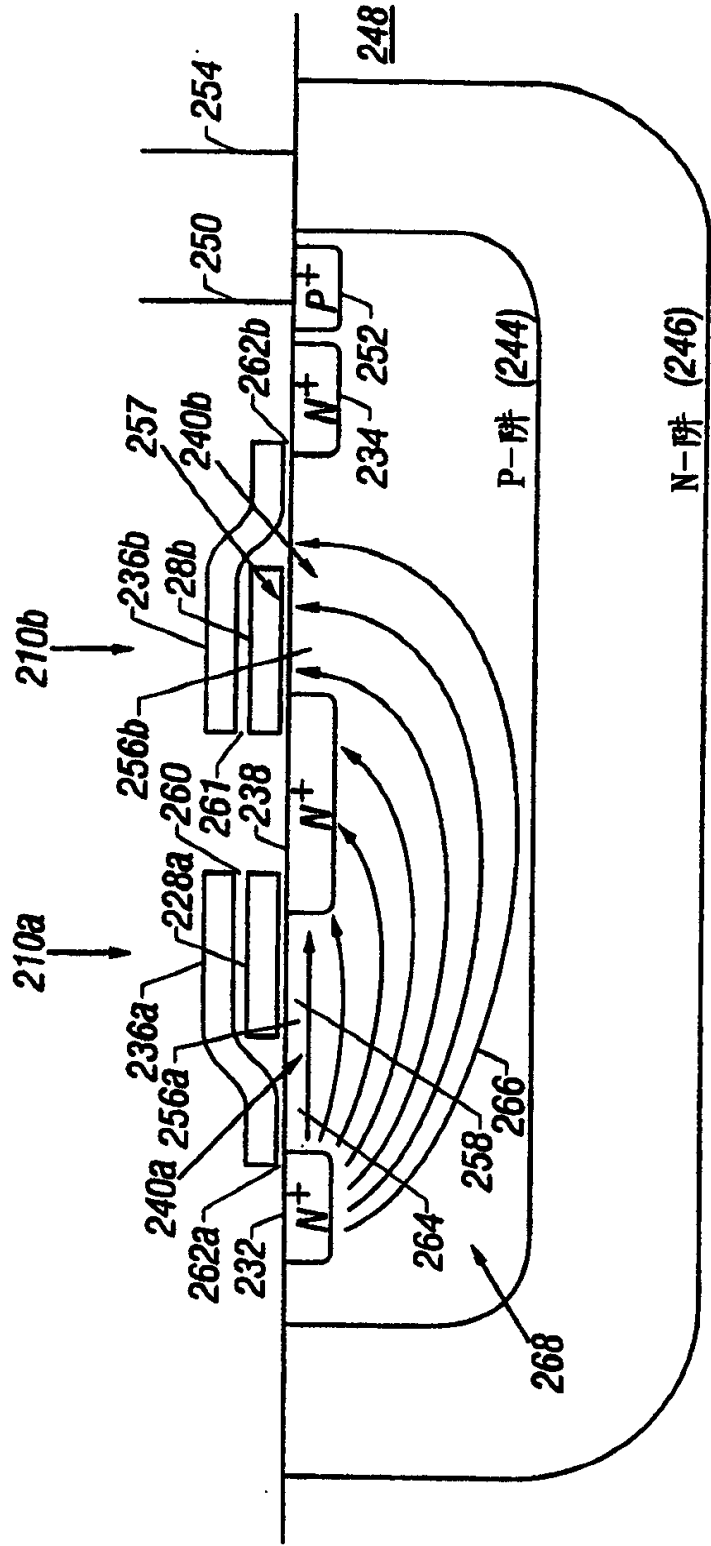


图 7

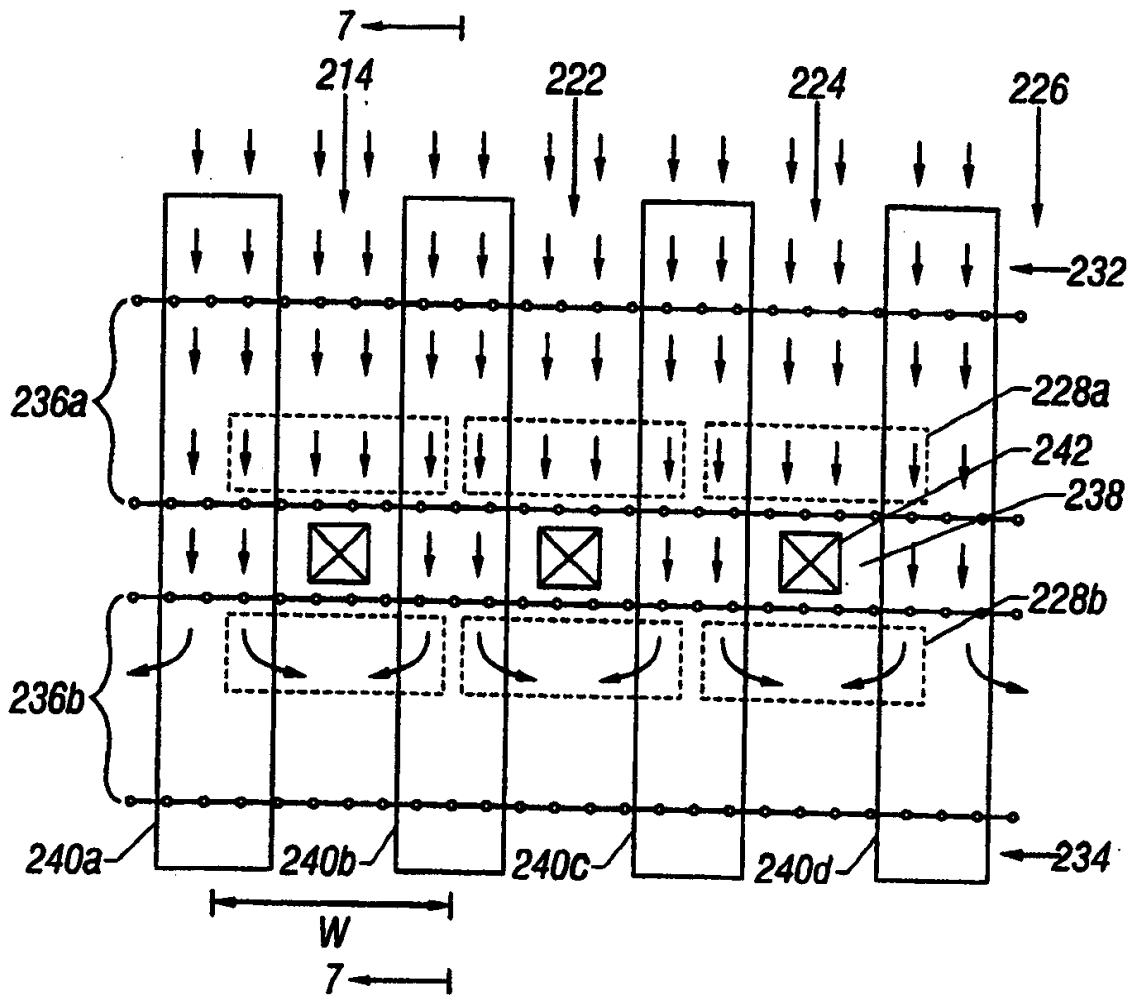


图 8

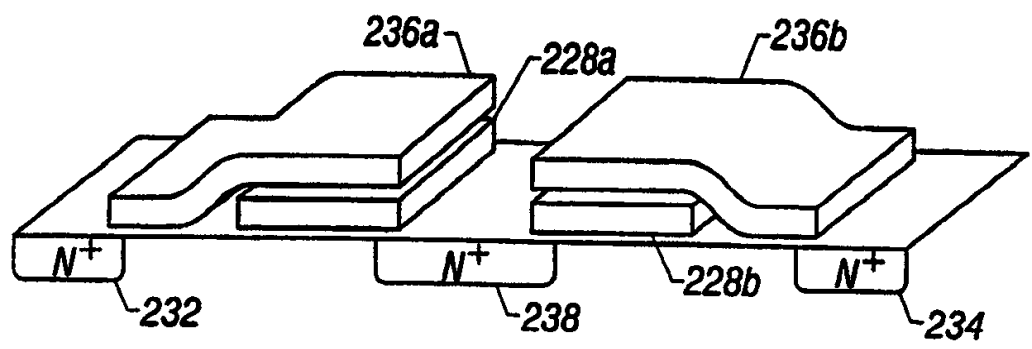


图 9

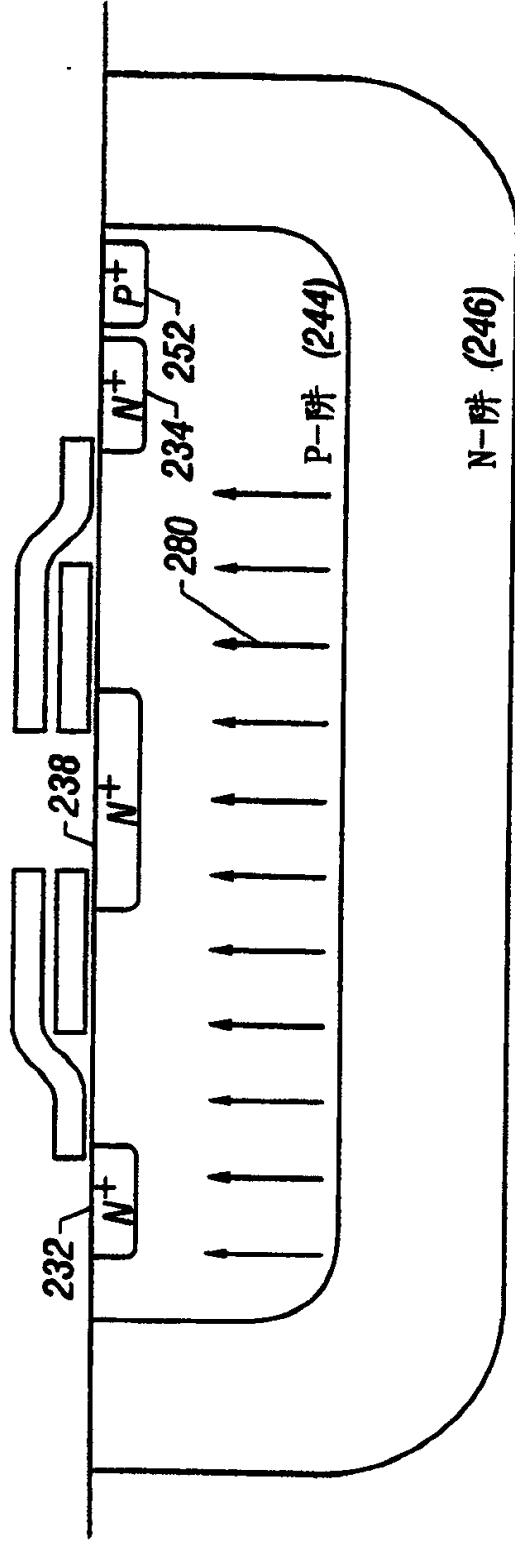


图 10