



(12) 发明专利申请

(10) 申请公布号 CN 117174138 A

(43) 申请公布日 2023.12.05

(21) 申请号 202310159975.8

(22) 申请日 2023.02.23

(30) 优先权数据

10-2022-0068464 2022.06.03 KR

(71) 申请人 三星电子株式会社

地址 韩国京畿道

申请人 高丽大学校 产学协力团

(72)发明人 朴钟善 李京镐 金弦峻

(74) 专利代理机构 中科专利商标代理有限责任

公司 11021

专利代理师 吴晓兵 倪斌

(51) Int.Cl.

G11C 11/412 (2006.01)

G11C 11/417 (2006.01)

G06F 7/544 (2006.01)

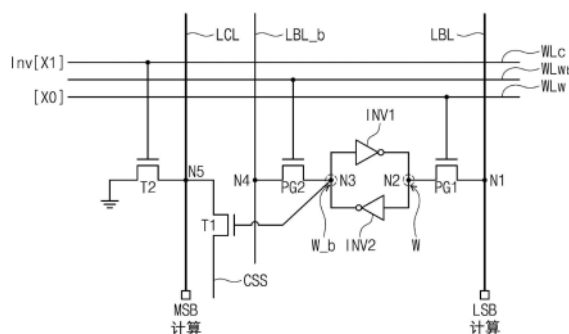
权利要求书3页 说明书15页 附图20页

(54) 发明名称

SRAM单元以及操作该SRAM单元的方法

(57) 摘要

SRAM单元包括:第一传输门晶体管,与第一字线和局部位线连接;第一反相器,包括与第一传输门晶体管连接的输出端、以及输入端;第二反相器,包括与第一传输门晶体管连接的输入端、以及输出端;第二传输门晶体管,与第二字线、第一反相器的输入端和第二反相器的输出端、以及互补局部位线连接;第一晶体管,与第二传输门晶体管、局部计算线和接地电极连接;以及第二晶体管,与第三字线、局部计算线和接地电极连接。



1. 一种静态随机存取存储器SRAM单元,包括:

第一传输门晶体管,包括与第一字线连接的栅电极、与局部位线连接的第一端、以及第二端;

第一反相器,包括与所述第一传输门晶体管的第二端连接的输出端、以及输入端;

第二反相器,包括与所述第一传输门晶体管的第二端连接的输入端、以及输出端;

第二传输门晶体管,包括与第二字线连接的栅电极、与所述第一反相器的输入端和所述第二反相器的输出端连接的第一端、以及与互补局部位线连接的第二端;

第一晶体管,包括与所述第二传输门晶体管的第一端连接的栅电极、与局部计算线连接的第一端、以及第二端;以及

第二晶体管,包括与第三字线连接的栅电极、与所述局部计算线连接的第一端、以及与接地电极连接的第二端。

2. 根据权利要求1所述的SRAM单元,其中,通过所述局部计算线、所述第一晶体管和所述第二晶体管的结的结电容是通过所述局部位线和所述第一传输门晶体管的结的结电容的两倍。

3. 根据权利要求1所述的SRAM单元,其中,所述第一反相器的输出端和所述第二反相器的输入端基于输入到所述第一字线的第一信号和输入到所述第二字线的第二信号来存储权重,并且

其中,所述第一反相器的输入端和所述第二反相器的输出端基于输入到所述第一字线的所述第一信号和输入到所述第二字线的所述第二信号来存储互补权重。

4. 根据权利要求3所述的SRAM单元,其中,输入数据的最低有效比特LSB被输入到所述第一字线,并且所述输入数据的最高有效比特MSB的反相值被输入到所述第三字线。

5. 根据权利要求4所述的SRAM单元,其中,响应于输入到所述第一字线的所述LSB,所述第一传输门晶体管将所述权重传送到所述局部位线或允许维持所述局部位线的初始值。

6. 根据权利要求4所述的SRAM单元,其中,响应于所述互补权重,所述第一晶体管将所述局部计算线与所述接地电极连接或允许维持所述局部计算线的初始值,并且

其中,响应于输入到所述第三字线的所述MSB的所述反相值,所述第二晶体管将所述局部计算线与所述接地电极连接或允许维持所述局部计算线的初始值。

7. 根据权利要求6所述的SRAM单元,还包括:

第一开关,被配置为将所述局部计算线与所述局部位线电连接;以及

第二开关,被配置为将电连接的所述局部计算线和所述局部位线与全局位线电连接。

8. 一种存储器设备,包括:

交叉耦合反相器;

第一传输门晶体管,被配置为响应于输入数据的最低有效比特LSB将存储在所述交叉耦合反相器的第一节点上的权重传送到局部位线;

第二传输门晶体管,被配置为将存储在所述交叉耦合反相器的第二节点上的互补权重传送到互补局部位线;

第一晶体管,被配置为响应于所述互补权重将存储在局部计算线上的电荷输出到接地计算线;

第二晶体管,被配置为响应于所述输入数据的最高有效比特MSB的反相值将存储在所

述局部计算线上的所述电荷放电;以及

读出电路,

其中,响应于所述互补局部位线的电压,所述读出电路将输出到所述接地计算线的所述电荷放电,并且将所述局部计算线与所述局部位线电连接。

9. 根据权利要求8所述的存储器设备,其中,所述读出电路包括:

第三晶体管,被配置为响应于所述互补局部位线的电压对输出到所述接地计算线的所述电荷进行传送;

第四晶体管,被配置为将由所述第三晶体管传送的所述电荷输出到接地电极;

第一开关,被配置为将所述局部计算线电连接到所述局部位线;以及

第二开关,被配置为将电连接的所述局部计算线和所述局部位线与全局位线电连接。

10. 根据权利要求8所述的存储器设备,还包括:

预充电电路,被配置为对所述局部计算线和所述局部位线进行预充电或放电。

11. 根据权利要求8所述的存储器设备,还包括:

行解码器,被配置为选择和驱动字线,所述输入数据的所述LSB和所述输入数据的所述MSB的所述反相值通过所述字线被传送;以及

列解码器,被配置为选择与彼此电连接的所述局部计算线和所述局部位线电连接的全局位线。

12. 根据权利要求11所述的存储器设备,还包括:

模数转换器,被配置为检测所述全局位线的电压电平,并且输出与所述电压电平相对应的值,其中通过所述局部计算线与所述局部位线的电连接所共享的电荷被输出到所述全局位线;以及

移位加法器,被配置为对从所述模数转换器输出的所述值执行移位加法。

13. 根据权利要求11所述的存储器设备,还包括:

控制逻辑电路,被配置为从外部接收命令和地址,并且基于所述命令和所述地址生成用于选择所述字线的行地址和用于选择所述全局位线的列地址。

14. 根据权利要求8所述的存储器设备,其中,通过所述局部计算线、所述第一晶体管和所述第二晶体管的结的结电容是通过所述局部位线和所述第一传输门晶体管的结的结电容的两倍。

15. 一种存储器设备,包括:

第一静态随机存取存储器SRAM单元,被配置为在第一节点上存储第一权重以及在第二节点上存储第一互补权重,其中所述第一SRAM单元包括:

第一传输门晶体管,被配置为响应于第一输入数据的最低有效比特LSB将所述第一权重输出到第一局部位线;

第一晶体管,被配置为响应于所述第一互补权重将第一局部计算线与接地电极连接;以及

第二晶体管,被配置为响应于所述第一输入数据的最高有效比特MSB的反相值将所述第一局部计算线与所述接地电极连接;第一读出电路,被配置为将所述第一局部计算线和所述第一局部位线与全局位线电连接;

第二SRAM单元,被配置为在第三节点上存储第二权重以及在第四节点上存储第二互补

权重,其中所述第二SRAM单元包括:

第三传输门晶体管,被配置为响应于第二输入数据的LSB将所述第二权重输出到第二局部位线;

第三晶体管,被配置为响应于所述第二互补权重将第二局部计算线与所述接地电极连接;以及

第四晶体管,被配置为响应于所述第二输入数据的MSB的反相值将所述第二局部计算线与所述接地电极连接;以及

第二读出电路,被配置为将所述第二局部计算线和所述第二局部位线与所述全局位线电连接。

16. 根据权利要求15所述的存储器设备,还包括:

预充电电路,被配置为对所述第一局部计算线和所述第二局部计算线以及所述第一局部位线和所述第二局部位线进行预充电或放电。

17. 根据权利要求15所述的存储器设备,还包括:

第一局部行解码器,被配置为选择和驱动第一字线,所述第一输入数据的所述LSB和所述第一输入数据的所述MSB的所述反相值通过所述第一字线被传送;

第二局部行解码器,被配置为选择和驱动第二字线,所述第二输入数据的所述LSB和所述第二输入数据的所述MSB的所述反相值通过所述第二字线被传送;以及

列解码器,被配置为选择所述全局位线。

18. 根据权利要求15所述的存储器设备,还包括:

模数转换器,被配置为检测所述全局位线的电压电平并且输出与所述全局位线的所述电压电平相对应的值;以及

移位加法器,被配置为对从所述模数转换器输出的所述值执行移位加法。

19. 根据权利要求15所述的存储器设备,还包括:

控制逻辑电路,被配置为从外部接收命令和地址,并且基于所述命令和所述地址生成用于选择所述第一字线和所述第二字线的行地址以及用于选择所述全局位线的列地址。

20. 根据权利要求15所述的存储器设备,其中,通过所述第一局部计算线、所述第一晶体管和所述第二晶体管的结的结电容是通过所述第一局部位线和所述第一传输门晶体管的结的结电容的两倍,并且

其中,通过所述第二局部计算线、所述第三晶体管和所述第四晶体管的结的结电容是通过所述第二局部位线和所述第三传输门晶体管的结的结电容的两倍。

SRAM单元以及操作该SRAM单元的方法

[0001] 相关申请的交叉引用

[0002] 本申请要求于2022年6月3日向韩国知识产权局递交的韩国专利申请No.10-2022-0068464的优先权,其整体公开一并于此以作参考。

技术领域

[0003] 本公开涉及存储器设备,并且更具体地,涉及基于电荷共享对多比特数据执行乘法累加(MAC)运算的静态随机存取存储器(SRAM)单元以及操作该SRAM单元的方法。

背景技术

[0004] 存储器和处理器在冯·诺依曼架构的设备中物理分离。由于存储数据的位置和执行操作的位置不同,因此在存储器与处理器之间交换数据时需要大量的能量(即大量的功率)。可以使用具有内存计算架构的设备来减少由于数据交换而导致的过度能量消耗。

[0005] 例如,具有内存计算架构的设备可以基于电流或使用诸如电容器之类的单独存储元件来执行MAC运算。然而,基于电流执行MAC运算的设备具有低线性特性(即,非线性)。此外,这些设备制造起来可能很昂贵,占用大量面积,并且无法执行准确的多比特运算。

[0006] 因此,需要开发一种能够以高精度和良好线性度执行多比特运算的设备。

发明内容

[0007] 在此描述的本公开的实施例涉及存储器设备,并且更具体地,涉及对多比特数据和权重执行乘法累加(MAC)运算的静态随机存取存储器(SRAM)单元。

[0008] 本公开的实施例提供一种被配置为基于电荷共享来执行MAC运算的SRAM单元。

[0009] 本公开的实施例提供一种被配置为在没有单独存储元件(比如电容器)的情况下执行MAC运算的SRAM单元。

[0010] 根据实施例,一种静态随机存取存储器(SRAM)单元包括:第一传输门晶体管,包括与第一字线连接的栅电极、与局部位线连接的第一端、以及第二端;第一反相器,包括与第一传输门晶体管的第二端连接的输出端、以及输入端;第二反相器,包括与第一传输门晶体管的第二端连接的输入端、以及输出端;第二传输门晶体管,包括与第二字线连接的栅电极、与第一反相器的输入端和第二反相器的输出端连接的第一端、以及与互补局部位线连接的第二端;第一晶体管,包括与第二传输门晶体管的第一端连接的栅电极、与局部计算线连接的第一端、以及与接地电极连接的第二端;以及第二晶体管,包括与第三字线连接的栅电极、与局部计算线连接的第一端、以及与接地电极连接的第二端。

[0011] 根据实施例,一种存储器设备包括:交叉耦合反相器;第一传输门晶体管,响应于输入数据的LSB将存储在交叉耦合反相器的第一节点上的权重传送到局部位线;第二传输门晶体管,将存储在交叉耦合反相器的第二节点上的互补权重传送到互补局部位线;第一晶体管,响应于互补权重将存储在局部计算线上的电荷输出到接地计算线;第二晶体管,响应于输入数据的MSB的反相值将存储在局部计算线上的电荷放电;以及读出电路。响应于互

补局部位线的电压,读出电路可以将输出到接地计算线的电荷放电,并且将局部计算线与局部位线电连接。

[0012] 根据实施例,一种存储器设备可以包括:第一SRAM单元,在第一节点上存储第一权重以及在第二节点上存储第一互补权重,第一SRAM单元包括:第一传输门晶体管,响应于第一输入数据的LSB将第一权重输出到第一局部位线;第一晶体管,响应于第一互补权重将第一局部计算线与接地电极连接;以及第二晶体管,响应于第一输入数据的MSB的反相值将第一局部计算线与接地电极连接;第一读出电路,将第一局部计算线和第一局部位线与全局位线电连接;第二SRAM单元,在第三节点上存储第二权重以及在第四节点上存储第二互补权重,第二SRAM单元包括:第三传输门晶体管,响应于第二输入数据的LSB将第二权重输出到第二局部位线;第三晶体管,响应于第二互补权重将第二局部计算线与接地电极连接;以及第四晶体管,响应于第二输入数据的MSB的反相值将第二局部计算线与接地电极连接;以及第二读出电路,将第二局部计算线和第二局部位线与全局位线电连接。

附图说明

[0013] 通过参照附图详细描述本公开的实施例,本公开的上述及其他目的和特征将变得显而易见。

[0014] 图1是示出了根据本公开的实施例的存储器设备的配置的图。

[0015] 图2是根据本公开的实施例的SRAM单元的电路图。

[0016] 图3A是示出了图2的局部位线与局部计算线之间的电荷共享的图。

[0017] 图3B是示出了由图2的SRAM单元执行的多比特MAC运算的表格。

[0018] 图4是根据本公开的实施例的SRAM单元和读出电路的电路图。

[0019] 图5A和图5B是用于描述图4的SRAM单元和读出电路的操作的时序图。

[0020] 图6A和图6B是用于描述图4的SRAM单元和读出电路的操作的时序图。

[0021] 图7A和图7B是用于描述图4的SRAM单元和读出电路的操作的时序图。

[0022] 图8A和图8B是用于描述图4的SRAM单元和读出电路的操作的时序图。

[0023] 图9示出了根据本公开的实施例的存储器设备的配置。

[0024] 图10是示出了图9的局部阵列的配置的图。

[0025] 图11是示出了如何使用图9的存储器设备来执行MAC运算的图。

[0026] 图12是示出了如何累加通过使用图11的存储器设备来执行MAC运算所得到的结果的图。

[0027] 图13是与存储器设备的一条全局位线连接的局部阵列和预充电电路的电路图。

[0028] 图14是示出了如何处理在全局位线上累加的MAC运算的结果的图。

[0029] 图15是示出了根据本公开的实施例的执行MAC运算的方法的流程图。

[0030] 图16是示出了应用了根据本公开的实施例的包括SRAM单元的存储器设备的系统的图。

具体实施方式

[0031] 下面,将以使本领域普通技术人员可以容易实现本发明的程度详细并清楚地描述本公开的实施例。

[0032] 在详细描述中,参考术语“单元”、“模块”、“...器”等描述的组件和附图中所示的功能块可以用软件、硬件、或其组合来实现。例如,软件可以是机器代码、固件、嵌入式代码和应用软件。例如,硬件可以包括电路、电子电路、处理器、计算机、集成电路、集成电路内核、压力传感器、惯性传感器、微机电系统(MEMS)、无源元件或者其组合。

[0033] 图1是示出了根据本公开的实施例的存储器设备100的配置的图。存储器设备100可以包括存储器单元阵列110、控制逻辑电路120、行解码器130(例如,解码器电路)、列解码器140(例如,解码器电路)、预充电电路150和处理电路160。

[0034] 存储器单元阵列110可以包括沿多行和多列布置的存储器单元。例如,存储器单元可以是静态随机存取存储器(SRAM),但本公开的实施例不限于此。构成存储器单元阵列110的存储器单元可以与字线WL和全局位线GBL连接。详细地,每个存储器单元可以通过局部计算线LCL和局部位线LBL与全局位线GBL连接。每个存储器单元可以存储通过一条或多条字线WL输入的至少一个比特(例如,权重)。通过其他字线WL,每个存储器单元可以接收输入数据的最高有效比特(MSB)的反相值(inverted value)以及输入数据的最低有效比特(LSB)。

[0035] 根据本公开的实施例的存储器单元可以对权重和由MSB和LSB组成的多比特输入数据执行乘法运算。可以将MSB与权重相乘的结果(以下称为MSB与权重的相乘结果)输出到局部计算线LCL,并且可以将LSB与权重相乘的结果(以下称为LSB与权重的相乘结果)输出到局部位线LBL。MSB与权重的相乘结果(具体地,与相乘结果相对应的电荷)可以通过局部计算线LCL与构成存储器单元的晶体管中的至少一个晶体管之间的结(junction),存储在具有结电容分量(junction capacitance component)的局部计算线LCL上。LSB与权重的相乘结果(具体地,与相乘结果相对应的电荷)可以通过局部位线LBL与构成存储器单元的晶体管中的至少另一晶体管之间的结,存储在具有结电容分量的局部位线LBL上。

[0036] 之后,当局部计算线LCL与局部位线LBL电连接在一起时,可以发生电荷共享,使得存储在局部计算线LCL和局部位线LBL上的总电荷可以对应于将权重与由MSB和LSB组成的多比特输入数据相乘的结果。存储在局部计算线LCL和局部位线LBL上的电荷可以通过全局位线GBL输出。

[0037] 控制逻辑电路120可以接收命令CMD、地址ADDR和控制信号CTRL。控制逻辑电路120可以基于命令CMD、地址ADDR和控制信号CTRL中的至少一个来产生行地址和列地址。控制逻辑电路120可以向行解码器130提供行地址,并且可以向列解码器140提供列地址。此外,控制逻辑电路120可以控制预充电电路150和处理电路160。

[0038] 行解码器130可以通过字线WL与存储器单元阵列110连接。行解码器130可以对从控制逻辑电路120接收的行地址进行解码以选择和驱动一条或多条字线WL。

[0039] 列解码器140可以对从控制逻辑电路120接收的列地址进行解码,并且可以选择全局位线GBL,使得由局部计算线LCL和局部位线LBL共享的电荷被输出。

[0040] 预充电电路150以预设电压电平对局部计算线LCL和局部位线LBL进行预充电,或者可以对局部计算线LCL和局部位线LBL进行放电以设置为接地电压。

[0041] 处理电路160可以对由列解码器140选择的列中的存储器单元之中的、由行解码器130选择并驱动的存储器单元执行读取操作或写入操作。处理电路160可以接收要被编程到所选择的存储器单元中的数据。例如,处理电路160可以包括用于执行写入操作的写入驱动器(例如,驱动器电路)。

[0042] 处理电路160可以对存储器单元阵列110执行读取操作。在读取操作中,处理电路160可以感测由列解码器140选择的全局位线GBL的电压电平,并且可以将所感测的电压电平(即,模拟信号)转换为数字信号。例如,转换后的数字信号可以包括2比特数据(即,权重与由MSB和LSB组成的输入数据的相乘结果)。处理电路160可以对转换后的数字信号执行移位相加,通过移位相加可以输出完整形式的读取数据“DATA”。

[0043] 处理电路160可以对存储器单元阵列110执行写入操作。在写入操作中,处理电路160可以将从外部接收的数据“DATA”写入由行解码器130和列解码器140选择的存储器单元。例如,处理电路160的写入驱动器可以执行写入操作。

[0044] 根据本公开的实施例的SRAM单元可以通过使用电荷共享对多比特数据和权重执行乘法运算。特别地,SRAM单元可以在其中存储相乘结果而无需诸如电容器之类的单独存储元件。这样,可以省略对数据执行计算的处理器与存储数据的存储器设备之间的数据传输,因此可以显著减少在数据传输期间使用的能量(即,功率)。此外,因为不需要诸如电容器之类的单独存储元件,所以可以提高SRAM单元的集成度。此外,因为使用了电荷共享,所以可以保证MAC结果的线性度。

[0045] 图2是根据本公开的实施例的SRAM单元的电路图。

[0046] SRAM单元可以包括第一晶体管T1、第二晶体管T2、第一传输门(pass gate)晶体管PG1、第二传输门晶体管PG2、以及两个PMOS晶体管(即,上拉晶体管)和两个NMOS晶体管(即,下拉晶体管)。PMOS晶体管之一和NMOS晶体管之一可以构成反相器INV1,并且PMOS晶体管中的另一个PMOS晶体管和NMOS晶体管中的另一个NMOS晶体管可以构成反相器INV2。也就是说,本公开的SRAM单元可以是8T SRAM单元。例如,反相器INV1可以由两个晶体管实现,并且反相器INV2可以由另外两个晶体管实现。反相器INV1和INV2可以称为交叉耦合反相器。

[0047] 第一传输门晶体管PG1可以包括与字线 WL_w 连接的栅电极、与局部位线LBL连接的第一端(即,节点N1)、以及与第一反相器INV1的输出端和第二反相器INV2的输入端连接的第二端(即,节点N2)。第二传输门晶体管PG2可以包括与字线 WL_b 连接的栅电极、与互补局部位线LBL_b连接的第一端(即,节点N4)、以及与第一反相器INV1的输入端和第二反相器INV2的输出端连接的第二端(即,节点N3)。第一晶体管T1可以包括与节点N3连接的栅电极、与局部计算线LCL连接的第一端、以及与接地计算线CSS连接的第二端。在实施例中,接地计算线CSS与接地电极连接或接收接地电压。第二晶体管T2可以包括与字线 WL_c 连接的栅电极、与局部计算线LCL连接的第一端(即,节点N5)、以及与接地电极连接的第二端。接地电极可以接收接地电压。

[0048] 在实施例中,局部位线LBL可以通过与第一传输门晶体管PG1的结而具有结电容分量。因此,当第一传输门晶体管PG1导通时,局部位线LBL可以存储在节点N2上存储的值(即,电荷)。局部计算线LCL可以通过与第一晶体管T1和第二晶体管T2的结而具有结电容分量。因此,当晶体管T1和T2未导通时,局部计算线LCL可以存储在节点N5上存储的值(即,电荷)。

[0049] 在实施例中,通过局部计算线LCL与晶体管T1和T2的结的结电容是通过局部位线LBL与晶体管PG1的结的结电容的两倍。在制造SRAM单元的工艺中可以考虑晶体管的尺寸以及局部位线LBL和局部计算线LCL中的每一个的宽度和厚度,使得局部计算线LCL的结电容是局部位线LBL的结电容的两倍。例如,晶体管的尺寸以及局部位线LBL和局部计算线LCL中的每一个的宽度和厚度可以具有使局部计算线LCL的结电容为局部位线LBL的结电容的两

倍的值。

[0050] 在实施例中,用于将权重“W”存储在SRAM单元中的编程操作可以在执行使用本公开的SRAM单元的多比特乘法累加(MAC)运算之前执行。参照图1和图2,全局位线GBL可以由列解码器140选择,字线 WL_{wb} 和 WL_w 可以由行解码器130选择,并且用于导通传输门晶体管PG1和PG2的电压值可以分别施加到字线 WL_{wb} 和 WL_w 。为了在节点N2上存储比特“1”,可以将比特“1”输入到与所选择的全局位线GBL连接的局部位线LBL,并且可以将比特“0”输入到与所选择的互补全局位线GBL_b连接的互补局部位线LBL_b。相反,为了在节点N2上存储比特“0”,可以将比特“0”输入到与所选择的全局位线GBL连接的局部位线LBL,并且可以将比特“1”输入到与所选择的互补全局位线GBL_b连接的互补局部位线LBL_b。

[0051] 在实施例中,为了使用根据本公开的实施例的SRAM单元来执行多比特MAC运算,可以将多比特输入数据的MSB的反相值(即, $Inv[X1]$) 和多比特输入数据的LSB(即, $[X0]$) 分别输入到字线 WL_c 和 WL_w 。例如,存储器单元阵列110或行解码器130可以包括用于将多比特输入数据的MSB反相的反相器。

[0052] 节点N5的逻辑值可以取决于局部计算线LCL的电压的初始值和多比特输入数据的MSB的反相值 $Inv[X1]$ 。另外,节点N5的逻辑值可以取决于局部计算线LCL的电压的初始值和节点N3的逻辑值(即,互补权重 W_b)。这是因为第一晶体管T1或第二晶体管T2的导通使局部计算线LCL放电,并且局部计算线LCL的电压的初始值决定了局部计算线LCL的电压。换句话说,与存储在局部计算线LCL上的电荷相对应的电压值可以指示对多比特输入数据的MSB和权重“W”执行乘法运算的结果。在实施例中,在使用局部计算线LCL执行乘法运算之前,预充电电路150可以对局部计算线LCL进行预充电。

[0053] 节点N1的逻辑值可以取决于节点N2的逻辑值(即,权重“W”)和多比特输入数据的LSB(即, $[X0]$)。换句话说,与存储在局部位线LBL上的电荷相对应的电压值可以指示对多比特输入数据的LSB和权重“W”执行乘法运算的结果。在实施例中,在通过使用局部位线LBL执行乘法运算之前,可以由预充电电路150对局部位线LBL进行预充电。

[0054] 根据以上描述,由SRAM单元执行的2比特输入数据与权重的乘法运算可以由以下等式1表示。其中, $iAct[1:0]$ 表示2比特输入数据, $iAct[1]$ 表示2比特输入数据的MSB, $iAct[0]$ 表示2比特输入数据的LSB,并且 $W[n]$ 表示权重。

$$[0055] \quad iAct[1:0] \times W[n] = \overline{(\overline{iAct[1]} + \overline{W[n]})} \times 2 + (iAct[0] \cdot W[n]) \quad [\text{等式 1}]$$

[0056] 在上面的等式1中, $\overline{(\overline{iAct[1]} + \overline{W[n]})}$ 表示对局部计算线LCL执行的MSB乘法运算(或计算),并且 $(iAct[0] \cdot W[n])$ 表示对局部位线LBL执行的LSB乘法运算(或计算)。另外,“2”乘以 $\overline{(\overline{iAct[1]} + \overline{W[n]})}$ 表示通过局部计算线LCL与晶体管T1、T2的结的结电容是通过局部位线LBL与晶体管PG1的结的结电容的2倍。例如,对输入数据“01”和权重“1”执行的乘法运算可以如下计算: $01 \times 1 = \left(\overline{0 + 1} \right) \times 2 + (1 \cdot 1) = 01$ 。

[0057] 在使用局部计算线LCL和局部位线LBL执行乘法运算以后,将局部计算线LCL与局部位线LBL彼此电性连接,并且因此局部计算线LCL与局部位线LBL之间发生电荷共享。多比

特输入数据的MSB与权重的相乘结果与多比特输入数据的LSB与权重的相乘结果之和可以通过电荷共享获得。这将在下面更全面地描述。

[0058] 图3A是示出了图2的局部位线LBL与局部计算线LCL之间的电荷共享的图。

[0059] 参照图3A,示出了如下示例:因为局部计算线LCL不存在电荷,局部计算线LCL的电压值为“0”(即,VSS),以及因为局部位线LBL上的电荷已充满,局部位线LBL的电压值为“1”(即,VDD)。当开关SW接通时,在局部计算线LCL与局部位线LBL之间发生电荷共享。根据参照图2描述的假设,当局部计算线LCL的电容是局部位线LBL的电容的两倍时,电荷共享后的电压值可以为“ $1/3 \times VDD$ ”,这对应于比特“01”。

[0060] 虽然未示出,但当局部计算线LCL的电压值为“0”(即,VSS)且局部位线LBL的电压值也为“0”(即,VSS)时,电荷共享后的电压值可以为“0”,这对应于比特“00”。当局部计算线LCL的电压值为“1”(即,VDD)且局部位线LBL的电压值也为“0”(即,VSS)时,电荷共享后的电压值可以为“ $2/3 \times VDD$ ”,这对应于比特“10”。当局部计算线LCL的电压值为“1”(即,VDD)且局部位线LBL的电压值也为“1”(即,VDD)时,电荷共享后的电压值可以为“VDD”,这对应于比特“11”。

[0061] 图3B是示出了由图2的SRAM单元执行的多比特MAC运算的表格。

[0062] 输入数据 $iAct[1:0]$ 可以由2比特组成,输入数据的MSB的反相值 $Inv[X1]$ 可以输入到字线 WL_C ,并且输入数据的LSB(即, $[X0]$)可以输入到字线 WL_W 。权重“W”可以由1比特组成;如参照图2所述,权重“W”可以在MAC运算之前预先存储在SRAM单元中。

[0063] 情况1示出了当权重“W”为“0”时,与使用局部计算线LCL的MSB乘法运算和使用局部位线LBL的LSB乘法运算相关联的值和电压。

[0064] 参照使用局部计算线LCL的MSB乘法运算,因为第一晶体管T1由存储在节点N3上的值(即“1”)导通,所以存储在具有结电容分量的局部计算线LCL上的电荷可以通过接地计算线CSS放电(即,可以将局部计算线LCL设置为“VSS”)。备选地,当输入数据的MSB为“0”时,第二晶体管T2可以由输入数据的MSB的反相值 $Inv[X1]$ 导通,并且因此存储在局部计算线LCL上的电荷可以通过第二晶体管T2放电(即,可以将局部计算线LCL设置为“VSS”)。

[0065] 参照使用局部位线LBL的LSB乘法运算,无论输入数据的LSB的值(即 $[X0]$)如何,局部位线LBL的电压值可以是“VSS”。原因如下。通过预充电电路150(参照图1)的放电操作,局部位线LBL的初始值可以是“0”;即使第一传输门晶体管PG1由输入到字线 WL_W 的LSB导通,并且存储在节点N2上的“0”输出到局部位线LBL。

[0066] 之后,当局部计算线LCL与局部位线LBL之间发生电荷共享时,局部计算线LCL和局部位线LBL的电压值可以是“VSS”。

[0067] 情况2示出了当权重“W”为“1”且输入数据为“00”时,与使用局部计算线LCL的MSB乘法运算和使用局部位线LBL的LSB乘法运算相关联的值和电压。

[0068] 参照使用局部计算线LCL的MSB乘法运算,因为作为输入数据的MSB的反相值 $Inv[X1]$ 的“1”通过字线 WL_C 输入到第二晶体管T2,所以第二晶体管T2导通。因此,局部计算线LCL可以放电到“VSS”(例如,接地电压)。参照使用局部位线LBL的LSB乘法运算,因为作为输入数据的LSB的值的“0”通过字线 WL_W 输入到第一传输门晶体管PG1,所以第一传输门晶体管PG1截止。因此,局部位线LBL可以将“VSS”维持在初始值。之后,当局部计算线LCL与局部位线LBL之间发生电荷共享时,局部计算线LCL和局部位线LBL的电压值可以是“VSS”。

[0069] 情况3示出了当权重“W”为“1”且输入数据为“01”时,与使用局部计算线LCL的MSB乘法运算和使用局部位线LBL的LSB乘法运算相关联的值和电压。

[0070] 参照使用局部计算线LCL的MSB乘法运算,因为作为输入数据的MSB的反相值Inv[X1]的“1”通过字线 WL_c 输入到第二晶体管T2,所以第二晶体管T2导通。因此,局部计算线LCL可以放电到“VSS”。参照使用局部位线LBL的LSB乘法运算,因为作为输入数据的LSB的值的“1”通过字线 WL_w 输入到第一传输门晶体管PG1,所以第一传输门晶体管PG1导通。因此,存储在节点N2上的“1”输出到局部位线LBL,并且因此局部位线LBL的电压为“VDD”。之后,当局部计算线LCL与局部位线LBL之间发生电荷共享时,因为局部计算线LCL的电容是局部位线LBL的电容的两倍,所以局部计算线LCL和局部位线LBL的电压可以是“ $1/3 \times VDD$ ”。

[0071] 情况4示出了当权重“W”为“1”且输入数据为“10”时,与使用局部计算线LCL的MSB乘法运算和使用局部位线LBL的LSB乘法运算相关联的值和电压。

[0072] 参照使用局部计算线LCL的MSB乘法运算,因为作为输入数据的MSB的反相值Inv[X1]的“0”通过字线 WL_c 输入到第二晶体管T2,所以第二晶体管T2截止。此外,第一晶体管T1通过存储在节点N3上的互补权重 W_b 而截止。因此,局部计算线LCL可以将“VDD”维持在初始值。参照使用局部位线LBL的LSB乘法运算,因为作为输入数据的LSB的值的“0”通过字线 WL_w 输入到第一传输门晶体管PG1,所以第一传输门晶体管PG1截止。因此,局部位线LBL可以将“VSS”维持在初始值。之后,当局部计算线LCL与局部位线LBL之间发生电荷共享时,局部计算线LCL和局部位线LBL的电压值可以是“ $2/3 \times VDD$ ”。

[0073] 情况5示出了当权重“W”为“1”且输入数据为“11”时,与使用局部计算线LCL的MSB乘法运算和使用局部位线LBL的LSB乘法运算相关联的值和电压。

[0074] 参照使用局部计算线LCL的MSB乘法运算,因为作为输入数据的MSB的反相值Inv[X1]的“0”通过字线 WL_c 输入到第二晶体管T2,所以第二晶体管T2截止。此外,第一晶体管T1通过存储在节点N3上的互补权重 W_b 截止。因此,局部计算线LCL可以将“VDD”维持在初始值。参照使用局部位线LBL的LSB乘法运算,因为作为输入数据的LSB的值的“1”通过字线 WL_w 输入到第一传输门晶体管PG1,所以第一传输门晶体管PG1导通。因此,存储在节点N2上的“1”输出到局部位线LBL,并且局部位线LBL的电压为“VDD”。之后,当局部计算线LCL与局部位线LBL之间进行电荷共享时,局部计算线LCL和局部位线LBL的电压可以是“VDD”。

[0075] 图4是根据本公开的实施例的SRAM单元和读出电路的电路图。图4的SRAM单元可以与图2的SRAM单元相同,并且读出电路可以输出由SRAM单元执行的MAC运算的结果。

[0076] 读出电路可以包括晶体管T3和T4以及开关SW1、SW2和SW3。晶体管T3可以响应于节点N4的电压而操作,并且晶体管T3的第一端可以与接地计算线CSS连接。晶体管T4可以响应于电压CSSE而操作,并且可以包括与晶体管T3的第二端连接的第一端以及与接地电极连接的第二端。

[0077] 开关SW1响应于信号eLOP将局部计算线LCL和局部位线LBL电连接。例如,当开关SW1由晶体管实现时,信号eLOP可以被提供到其栅电极。在这种情况下,局部计算线LCL与局部位线LBL之间可以发生电荷共享,并且因此MSB与权重的相乘结果和LSB与权重的相乘结果可以相加。

[0078] 开关SW2可以响应于信号eLBL将电荷共享结果(即,MSB与权重的相乘结果和LSB与权重的相乘结果之和)输出到全局位线GBL。例如,当开关SW2由晶体管实现时,信号eLBL可

以被提供到其栅电极。

[0079] 开关SW3可以响应于信号eLBL_b将互补局部位线LBL_b和互补全局位线GBL_b电连接,并且可以将存储在节点N3上的互补权重W_b输出到互补全局位线GBL_b。例如,当开关SW3由晶体管实现时,信号eLBL_b可以被提供到其栅电极。

[0080] 例如,开关SW1、SW2和SW3中的每一个可以由至少一个PMOS晶体管和/或至少一个NMOS晶体管组成。例如,控制开关SW1、SW2和SW3的信号eL0P、eLBL和eLBL_b可以由控制逻辑电路120(参见图1)和/或行解码器130(参见图1)产生。开关SW1和SW2在图4中被示为读出电路的组件,但可以被视为SRAM单元的组件。

[0081] 图5A和图5B是用于描述图4的SRAM单元和读出电路的操作的时序图。详细地,图5A示出了与输入数据为“00”且权重为“1”的情况相关联的时序图,并且图5B示出了与输入数据为“00”且权重为“0”的情况相关联的时序图。

[0082] 参照图1、图4和图5A,在t0到t3的时间段期间,可以执行SRAM单元的初始化;在t3到t5的时间段期间,可以执行乘法运算,并且在t5之后可以通过电荷共享将乘法运算的结果相加。

[0083] 首先,为了将局部计算线LCL、局部位线LBL和互补局部位线LBL_b初始化,在t0到t1的时间段期间,可以激活信号eL0P和eLBL_b,使得开关SW1和SW3接通。在从t0到t3的时间段期间,可以激活信号eLBL使得开关SW2接通。

[0084] 在t1到t3的时间段期间,预充电电路150可以对局部计算线LCL进行预充电,并且预充电电路150可以对局部位线LBL进行放电。根据以上描述,用于使用局部计算线LCL的MSB乘法运算的初始值(即,“1”)可以存储在局部计算线LCL上,并且用于使用局部位线LBL的LSB乘法运算的初始值(即,“0”)可以存储在局部位线LBL上。

[0085] 在从t2到t3的时间段期间,当“1”被输入到字线WL_{wb}时,第二传输门晶体管PG2可以导通,并且因此,存储在节点N3上的互补权重W_b(即,“0”)可以被输出到互补局部位线LBL_b。在这种情况下,晶体管T3可以通过互补权重W_b(即,“0”)截止。

[0086] 在t3到t4时间段期间,当输入数据的MSB的反相值(即,Inv[X1] = “1”)通过字线WL_c输入到第二晶体管T2时,第二晶体管T2可以导通。在这种情况下,当局部计算线LCL放电时,局部计算线LCL的电压可以从“1”转变为“0”。如参照图2和图3B所描述的,在t3到t4时间段期间执行的MSB乘法运算的结果可以存储在局部计算线LCL上,并且LSB乘法运算的结果可以存储在局部位线LBL上。

[0087] 在t4到t5的时间段期间,由于信号eL0P被激活,开关SW1接通,并且因此局部计算线LCL与局部位线LBL之间发生电荷共享。因为通过局部计算线LCL与晶体管T1和T2的结的结电容是通过局部位线LBL与第一传输门晶体管PG1的结的结电容的两倍,所以共享电荷量可以是MSB与权重的相乘结果和LSB与权重的相乘结果之和。

[0088] 之后,在t5之后的时间段内,可以激活信号eLBL,并且因此共享电荷可以被输出到全局位线GBL。

[0089] 除了互补局部位线LBL_b的波形之外,图5B的时序图与图5A的时序图相同,并且因此将省略额外的描述以避免冗余。

[0090] 参照不同于图5A的波形的互补局部位线LBL_b的波形,因为存储在节点N3上的互补权重W_b是“1”,所以互补局部位线LBL_b的电压可以在t0之后的整个时间段内维持“1”。

因此,在从 t_2 到 t_3 的时间段期间,第二传输门晶体管PG2被输入到字线 WL_{wb} 的信号导通;当在 t_3 激活信号CSSE时,晶体管T4导通,并且因此局部计算线LCL从 t_3 开始放电。在这种情况下,局部计算线LCL的电压可以从“1”转变为“0”。

[0091] 图6A和图6B是用于描述图4的SRAM单元和读出电路的操作的时序图。详细地,图6A示出了与输入数据为“01”且权重为“1”的情况相关联的时序图,并且图6B示出了与输入数据为“01”且权重为“0”的情况相关联的时序图。

[0092] 首先,参照图6A,因为输入数据的MSB为“0”并且权重“W”为“1”,所以使用局部计算线LCL的MSB乘法运算与参照图5A描述的相同,并且因此将省略额外的描述以避免冗余。

[0093] 在从 t_3 到 t_4 的时间段期间,作为输入数据的LSB(即, $X[0]$)的“1”通过字线 WL_w 输入到第一传输门晶体管PG1。在这种情况下,作为存储在节点N2上的权重“W”的“1”可以输出到局部位线LBL,并且因此局部位线LBL的电压电平可以设置为“1”。

[0094] 之后,在信号eLOP被激活的 t_4 处,局部计算线LCL和局部位线LBL彼此电连接,并且因此局部计算线LCL和局部位线LBL的电压电平可以为“ $1/3 \times VDD$ ”。

[0095] 参照图6B,由于存储在节点N3上的互补权重 W_b 为“1”,所以互补局部位线LBL_b的电压可以在 t_0 之后的整个时间段内维持“1”。因此,当第二传输门晶体管PG2和晶体管T4导通时,局部计算线LCL从 t_3 开始放电,并且因此局部计算线LCL的电压可以从“1”转变为“0”。

[0096] 在从 t_3 到 t_4 的时间段期间,作为输入数据的LSB(即, $X[0]$)的“1”通过字线 WL_w 输入到第一传输门晶体管PG1。在这种情况下,作为存储在节点N2上的权重“W”的“0”可以输出到局部位线LBL,并且因此局部位线LBL的电压电平可以设置为“0”。因此,即使在 t_4 处局部计算线LCL与局部位线LBL之间发生电荷共享,局部计算线LCL和局部位线LBL的电压值也可以维持“0”。

[0097] 图7A和图7B是用于描述图4的SRAM单元和读出电路的操作的时序图。详细地,图7A示出了与输入数据为“10”且权重为“1”的情况相关联的时序图,并且图7B示出了与输入数据为“10”且权重为“0”的情况相关联的时序图。图7A和图7B的波形与参照图5A、图5B、图6A和图6B描述的波形部分相似,并且因此将省略附加描述以避免冗余。

[0098] 因为输入数据的MSB的反相值(即, $Inv[X_1]$)为“0”,所以第二晶体管T2在整个时间段内截止。此外,由于存储在节点N3上的互补权重 W_b 为“0”,所以即使第二传输门晶体管PG2在从 t_2 到 t_3 的时间段导通,晶体管T3也不导通。在这种情况下,在 t_1 到 t_4 的时间段期间,局部计算线LCL的电压电平可以维持作为初始值的“1”。因为作为输入数据的LSB(即, $X[0]$)的“0”在整个时间段内被输入到字线 WL_w ,所以第一传输门晶体管PG1截止,并且因此局部位线LBL的电压电平维持作为初始值的“0”。

[0099] 之后,在信号eLOP被激活的 t_4 处,局部计算线LCL和局部位线LBL电连接在一起,并且因此局部计算线LCL和局部位线LBL的电压电平可以为“ $2/3 \times VDD$ ”。

[0100] 参照图7B,在 t_3 到 t_4 的时间段期间,由于输入数据的MSB的反相值(即, $Inv[X_1]$)为“0”,所以第二晶体管T2在整个时间段内截止。此外,由于存储在节点N3上的互补权重 W_b 为“0”,所以即使第二传输门晶体管PG2在从 t_2 到 t_3 的时间段期间导通,晶体管T3也不导通。在这种情况下,在 t_1 到 t_4 的时间段期间,局部计算线LCL的电压电平可以维持作为初始值的“1”。

[0101] 在 t_2 到 t_3 的时间段期间,第二传输门晶体管PG2由输入到字线 WL_{wb} 的信号导通,并

且第一晶体管T1由作为存储在节点N3上的互补权重 W_b 的“1”导通。在 t_3 到 t_4 的时间段期间,晶体管T4可以通过信号CSSE导通,并且因此局部计算线LCL的电压电平可以设置为“0”。

[0102] 在局部位线LBL的情况下,因为输入数据的LSB(即, $[X_0]$)为“0”,所以第一传输门晶体管PG1可以在整个时间段内截止,并且因此局部位线LBL的电压电平可以维持“0”。

[0103] 因此,即使在 t_4 处局部计算线LCL与局部位线LBL之间发生电荷共享,局部计算线LCL和局部位线LBL的电压值也可以维持“0”。

[0104] 图8A和图8B是用于描述图4的SRAM单元和读出电路的操作的时序图。详细地,图8A示出了与输入数据为“11”且权重为“1”的情况相关联的时序图,并且图8B示出了与输入数据为“11”且权重为“0”的情况相关联的时序图。

[0105] 参照图8A,因为输入数据的MSB的反相值(即, $Inv[X_1]$)为“0”,所以第二晶体管T2在整个时间段内截止。此外,由于存储在节点N3上的互补权重 W_b 为“0”,所以即使第二传输门晶体管PG2在从 t_2 到 t_3 的时间段期间导通,晶体管T3也不导通。在这种情况下,在 t_1 到 t_4 的时间段期间,局部计算线LCL的电压电平可以维持作为初始值的“1”。

[0106] 在从 t_3 到 t_4 的时间段期间,作为输入数据的LSB(即, $[X_0]$)的“1”输入到字线 WL_w ,并且因此第一传输门晶体管PG1导通。在这种情况下,作为存储在节点N1上的权重“W”的“1”输出到局部位线LBL,并且因此局部位线LBL的电压电平可以从“0”转变为“1”。

[0107] 之后,在信号eL0P被激活的 t_4 处,当局部计算线LCL与局部位线LBL之间发生电荷共享时,局部计算线LCL和局部位线LBL的电压值可以维持“1”。

[0108] 参照图8B,即使局部计算线LCL的电压值在从 t_1 到 t_4 的时间段期间维持在“1”,但是在 t_3 处,由于作为存储在节点N3上的互补权重 W_b 的“1”以及信号CSSE的激活,局部计算线LCL的电压值从“1”转变为“0”。在 t_3 到 t_4 的时间段期间,由于输入数据的LSB(即, $[X_0]$)被输入到字线 WL_w ,作为存储在节点N3上的权重“W”的“0”输出到局部位线LBL。因此,局部位线LBL的电压值可以维持“0”。之后,在信号eL0P被激活的 t_4 处,当局部计算线LCL与局部位线LBL之间发生电荷共享时,局部计算线LCL和局部位线LBL的电压值可以维持“0”。

[0109] 图9是示出了根据本公开的实施例的存储器设备200的配置的图。图10是示出了图9的局部阵列的配置的图。

[0110] 参照图9和图10,存储器设备200可以包括存储器单元阵列210、控制逻辑电路220、行解码器230、列解码器240、预充电电路250和处理电路260。图9的存储器设备200大部分类似于图1的存储器设备100,并且因此将省略额外的描述以避免冗余。

[0111] 存储器单元阵列210可以包括多个局部阵列。在实施例中,局部阵列被布置成矩形的形状(例如, 16×8 的维度)。每个局部阵列可以包括多个SRAM单元和读出电路。图10所示的每个SRAM单元可以与参照图2描述的SRAM单元相同,并且读出电路可以与图4的读出电路相同。在实施例中,局部阵列可以包括16个SRAM单元,并且包括在一个局部阵列中的SRAM单元可以共享局部位线LBL、局部计算线LCL、互补局部位线LBL_b、全局位线GBL和互补全局位线GBL_b。然而,包括在一个局部阵列中的SRAM单元不需要共享字线。也就是说,字线 WL_c 、 WL_{w0} 和 WB_w 可以与包括在一个局部阵列中的每个SRAM单元连接。

[0112] 行解码器230可以包括多个局部行解码器。在实施例中,如图9所示存在16个局部行解码器230_1至230_16。每个局部行解码器可以选择和驱动针对每一行设置的局部阵列的SRAM单元。例如,第一局部行解码器230_1可以选择并驱动与第一字线 $WL_s[1]$ 连接的SRAM

单元。如图2所示,第一字线 $W_{Ls}[1]$ 可以包括字线 W_{Lc} 、 W_{Lwb} 和 W_{Lw} 。第一局部行解码器230_1可以通过字线 W_{Lc} 和 W_{Lw} 向局部阵列提供2比特输入数据 $iAct_01[1:0]$ 。

[0113] 列解码器240可以选择包括所选择的SRAM单元的局部阵列的全局位线对GBL和GBL_b。包括在局部阵列中的SRAM单元可以共享全局位线对GBL和GBL_b。

[0114] 图11是示出了如何使用图9的存储器设备来执行MAC运算的图。为了附图简洁,仅示出了构成存储器单元阵列210的局部阵列,将参照图9至图11进行描述。

[0115] 存储器设备200可以作为对输入数据和权重执行乘法运算的人工神经网络来操作。例如,权重可以与人工神经网络的节点相关联,并且乘法运算可以是该节点的运算。在执行MAC运算之前,可以执行用于将权重存储在SRAM单元中的编程操作。例如,局部阵列[1]的SRAM单元中执行MAC运算的SRAM单元可以存储权重“1”;局部阵列[2]的SRAM单元中执行MAC运算的SRAM单元可以存储权重“1”;并且局部阵列[3]的SRAM单元中执行MAC运算的SRAM单元可以存储权重“0”。如以上描述,局部阵列[8]的SRAM单元中执行MAC运算的SRAM单元可以存储权重“1”。下面,为了便于描述,假设权重存储在与参照图4描述的字线 W_{Lc} 、 W_{Lwb} 和 W_{Lw} 共同连接的SRAM单元中。

[0116] 行解码器230可以向存储器单元阵列210提供输入数据。例如,第一局部行解码器230_1可以通过第一字线 $W_{Ls}[1]$ 向局部阵列[1]至局部阵列[8]提供2比特输入数据 $iAct_01[1:0]$ 。详细地,第一局部行解码器230_1可以向局部阵列[1]的第一SRAM单元提供输入数据“10”。输入数据“10”可以用灰色阴影标记。第一局部行解码器230_1可以向局部阵列[2]的第一SRAM单元提供输入数据“10”,并且可以向局部阵列[3]的第一SRAM单元提供输入数据“11”。如以上描述,第一局部行解码器230_1可以向局部阵列[8]的第一SRAM单元提供输入数据“10”。

[0117] 当输入数据 $iAct_01[1:0]$ (即,“10”)通过第一字线 $W_{Ls}[1]$ 传送时,可以对输入数据 $iAct_01[1:0]$ 和存储在每个SRAM单元中的权重执行部分乘积。例如,局部阵列[1]的第一SRAM单元可以对输入数据“10”(具体地,“00”,其中输入数据“10”的MSB被反相)和权重“1”执行乘法运算作为第一部分乘积PP1_R1;局部阵列[2]的第一SRAM单元可以对输入数据“10”和权重“1”执行乘法运算作为第二部分乘积PP2_R1;局部阵列[3]的第一SRAM单元可以对输入数据“10”和权重“0”执行乘法运算作为第三部分乘积PP3_R1。如以上描述,局部阵列[8]的第一SRAM单元可以对输入数据“10”和权重“1”执行乘法运算作为第八部分乘积PP8_R1。根据以上部分乘积,局部阵列[1]至局部阵列[8]的SRAM单元可以通过全局位线GBL[1]至GBL[8]输出作为部分乘积结果的“10”、“10”、“11”、“10”、“10”、“00”、“10”、“10”和“10”。

[0118] 同时,在与一个局部行解码器连接的局部阵列中执行的MAC运算可以在与另一局部行解码器连接的局部阵列中以相同的方式执行。例如,与第二局部行解码器230_2连接的局部阵列[1]至局部阵列[8]可以通过全局位线GBL[1]至GBL[8]分别输出作为部分乘积结果的“00”、“00”、……、“11”。另外,与第十六局部行解码器230_16连接的局部阵列[1]至局部阵列[8]可以通过全局位线GBL[1]至GBL[8]可以分别输出作为部分乘积结果的“00”、“00”、……、“01”。

[0119] 图12是示出了如何累加通过使用图11的存储器设备来执行MAC运算所得到的结果的图。

[0120] 在实施例中,部分乘积结果可以在相同的时刻处从每一行输出。例如,当需要将部

分乘积结果相加时,可以在与局部行解码器230_1至230_16连接的局部阵列中同时执行MAC运算。在与一条全局位线连接的局部阵列中执行的MAC运算的结果可以同时输出到全局位线。例如,部分乘积结果PP1_R1、PP1_R2、...、PP1_R16可以从与第一全局位线GBL[1]连接的局部阵列输出,并且与部分乘积结果PP1_R1、PP1_R2、...、PP1_R16中的每一个相对应的电荷可以在第一全局位线GBL[1]上累加。如以上描述,在其余局部阵列中执行的部分乘积的结果可以输出到全局位线GBL[2]至GBL[16]。

[0121] 在实施例中,部分乘积结果可以在每行的不同时刻处输出。例如,当不需要相加部分乘积结果时,不需要同时输出部分乘积结果。例如,在与第一局部行解码器230_1连接的局部阵列中执行MAC运算之后,可以在与第二局部行解码器230_2连接的局部阵列中执行MAC运算。

[0122] 图13是与存储器设备200的一条全局位线连接的局部阵列和预充电电路的电路图。

[0123] 参照图4、图9和图13,在计算全局位线GBL的电压之前,可以如图13所示对与一条全局位线GBL连接的组件进行建模。例如,一个局部阵列可以用“3Cp”表示。在图13中,Cp可以表示通过局部位线LBL与第一传输门晶体管PG1的结的结电容,并且2Cp可以表示通过局部计算线LCL与晶体管T1和T2的结的结电容。在图13中,C_{GBL}可以表示全局位线GBL的电容,并且C_{SH}可以表示列解码器240的输出开关控制信号SWCTRL的输出端的任何其他电容分量。

[0124] 根据电荷量与电容之间的关系公式,全局位线GBL的电压可以由以下等式2表示。在下面的等式2中,ΔV表示每个局部阵列的电压变化。

$$[0125] \quad V_{GBL} = \frac{\sum 3Cp \times \Delta V}{16 \times 3Cp + C_{GBL} + C_{SH}} \quad [\text{等式 2}]$$

[0126] 根据上面的等式2,因为局部阵列的电容之和大于剩余电容C_{GBL}和C_{SH},所以可以认为根据MAC运算结果的值的全局位线GBL的电压V_{GBL}的线性度被维持。根据本公开的存储器设备可以通过使用局部计算线LCL与局部位线LBL之间的电荷共享来执行MAC运算,并且因此可以获得线性度接近于实际理想值的模拟MAC电压。存储器设备200可以包括具有接收信号PCH的栅极的第一晶体管以及具有接收信号DCH的栅极的第二晶体管。

[0127] 图14是示出了如何处理在全局位线上累加的MAC运算的结果的图。

[0128] 参照图9、图11和图14,处理电路160可以包括模数转换器242和移位加法器244。

[0129] 在与第一全局位线GBL[1]连接的局部阵列中执行的MAC运算的结果PR1_R1、PR1_R2和PR1_R16可以在第一全局位线GBL[1]上累加以产生累加结果,并且累加结果可以输入到ADC 242。ADC 242可以基于累加结果输出第一部分MAC值pMAC[1]。如以上描述,ADC242可以输出第七部分MAC值pMAC[7]和第八部分MAC值pMAC[8]。

[0130] 移位加法器244可以接收第一部分MAC值pMAC[1]至第八部分MAC值pMAC[8]。移位加法器244可以基于第一部分MAC值pMAC[1]至第八部分MAC值pMAC[8]输出完整形式的MAC运算结果。例如,移位加法器244可以将2⁰与第一部分MAC值pMAC[1]相乘,可以将2⁶与第七部分MAC值pMAC[7]相乘,并且可以将(-2)⁷与第八部分MAC值pMAC[8]相乘,并将相乘结果相加。这里,(-2)⁷可以乘以部分MAC值来表示负的MAC运算结果。

[0131] 图15是示出了根据本公开的实施例的执行MAC运算的方法的流程图。

[0132] 参照图1、图4和图15,在操作S110中,对SRAM单元执行编程操作。例如,编程操作可

以将权重存储在SRAM单元中。权重“W”可以通过字线 WL_{wb} 和 WL_w 输入到SRAM单元,并且权重“W”和互补权重 W_b 可以分别存储在节点N2和节点N3上。

[0133] 在操作S120中,SRAM单元接收多比特输入数据。

[0134] 在操作S130中,SRAM单元通过使用响应于通过字线 WL_w 输入的输入数据的LSB操作的第一传输门晶体管PG1、存储在节点N2上的权重“W”、以及局部位线LBL的电压的初始值来确定局部位线LBL的电压电平。

[0135] 在操作S140中,SRAM单元通过使用第一晶体管T1、第二晶体管T2、以及局部计算线LCL的电压的初始值来确定局部计算线LCL的电压电平,该第一晶体管T1响应于存储在SRAM单元中的互补权重 W_b 而操作,该第二晶体管T2响应于通过字线 WL_c 输入的输入数据的MSB的反相值而操作。在实施例中,局部计算线LCL的电压电平的确定和局部位线LBL的电压电平的确定可以同时进行,或者可以不同时进行。

[0136] 在实施例中,在SRAM单元中执行MAC运算之前,预充电电路150可以对局部计算线LCL进行预充电并且可以对局部位线LBL进行放电。

[0137] 在操作S150中,将局部计算线LCL与局部位线LBL电连接,使得局部计算线LCL和局部位线LBL的电荷共享。例如,局部位线LBL与局部计算线LCL可以电连接在一起。之后,可以将局部计算线LCL和局部位线LBL与全局位线GBL电连接,并且因此共享电荷可以通过全局位线GBL输出。

[0138] 在操作S160中,确定与共享电荷相对应的电压电平。例如,处理电路160的模数转换器可以检测全局位线GBL的电压电平并且可以输出与由此检测到的电压电平相对应的MAC值。

[0139] 图16是根据实施例的应用了包括SRAM的存储器设备的系统1000的图。图16的系统1000可以是移动系统,例如便携式通信终端(例如,移动电话)、智能手机、平板个人计算机(PC)、可穿戴设备、保健设备、互联网、或物联网(IOT)设备。然而,图16的系统1000不一定限于移动系统,并且可以是PC、膝上型计算机、服务器、媒体播放器或汽车设备(例如,导航设备)。

[0140] 参照图16,系统1000可以包括主处理器1100、存储器(例如,1200a和1200b)和存储设备(例如,1300a和1300b)。此外,系统1000可以包括图像捕获设备1410、用户输入设备1420、传感器1430、通信设备1440、显示器1450、扬声器1460、供电设备1470和连接接口1480中的至少一个。

[0141] 主处理器1100可以控制系统1000的所有操作,更具体地,包括在系统1000中的其他组件的操作。主处理器1100可以被实施为通用处理器、专用处理器或应用处理器。

[0142] 主处理器1100可以包括至少一个CPU核1110,还包括控制器1120,该控制器1120被配置为控制存储器1200a和1200b和/或存储设备1300a和1300b。在一些实施例中,主处理器1100还可以包括加速器1130,该加速器1130是用于高速数据运算(诸如人工智能(AI)数据运算)的专用电路。加速器1130可以包括图形处理单元(GPU)、神经处理单元(NPU)和/或数据处理单元(DPU),并且被实现为与主处理器1100的其他组件在物理上分离的芯片。

[0143] 存储器1200a和1200b可以用作系统1000的主存储器设备。尽管存储器1200a和1200b中的每一个都可以包括易失性存储器,诸如静态随机存取存储器(SRAM)和/或动态RAM(DRAM),但是存储器1200a和1200b中的每一个都可以包括非易失性存储器,诸如闪存、

相变RAM (PRAM) 和/或电阻式RAM (RRAM)。存储器1200a和1200b可以在与主处理器1100相同的封装中实现。

[0144] 在实施例中,存储器1200a和1200b中的至少一个可以包括参照图1到图15描述的SRAM单元。存储器1200a和1200b中的至少一个可以作为对多比特输入数据和权重执行MAC运算的人工神经网络来操作。如果主处理器1100与存储器设备之间的数据传输是针对MAC运算而执行的,那么这将导致功耗增加和计算效率降低。然而,根据本公开的实施例,MAC运算完全由存储器1200a和1200b中的至少一个执行。由此,可以防止根据数据传输的功耗,并且可以提高计算效率。

[0145] 存储器设备1300a和1300b可以用作非易失性存储设备,并且具有比存储器1200a和1200b大的存储容量,该非易失性存储设备被配置为无论是否向其供电都存储数据。存储设备1300a和1300b可以分别包括存储控制器 (STRG CTRL) 1310a和1310b、以及被配置为经由存储控制器1310a和1310b的控制来存储数据的NVM (非易失性存储器) 1320a和1320b。虽然NVM 1320a和1320b可以包括具有二维 (2D) 结构或三维 (3D) V-NAND结构的闪存,但是NVM 1320a和1320b可以包括其他类型的NVM,诸如PRAM和/或PRAM。

[0146] 存储设备1300a和1300b可以在物理上与主处理器1100分离并且包括在系统1000中或者在与主处理器1100相同的封装中实现。此外,存储设备1300a和1300b可以具有固态设备 (SSD) 或存储卡的类型,并且可以通过接口 (诸如将在下文描述的连接接口1480) 与系统100的其它组件可移除地组合。存储设备1300a和1300b可以是应用了标准协议的设备,诸如通用闪存 (UFS)、嵌入式多媒体卡 (eMMC) 或非易失性存储器快速 (NVMe),并不限于此。

[0147] 图像捕获设备1410可以捕获静止图像或运动图像。图像捕获设备1410可以包括相机、摄像机和/或网络摄像头。

[0148] 用户输入设备1420可以接收由系统1000的用户输入的各种类型的数据,并且包括触摸板、键区、键盘、鼠标和/或麦克风。

[0149] 传感器1430可以检测可以从系统1000外部获得的各种类型的物理量,并将检测到的物理量转换为电信号。传感器1430可以包括温度传感器、压力传感器、照度传感器、位置传感器、加速度传感器、生物传感器和/或陀螺仪传感器。

[0150] 通信设备1440可以根据各种通信协议在系统1000外部的其他设备之间发送和接收信号。通信设备1440可以包括天线、收发器和/或调制解调器。

[0151] 显示器1450和扬声器1460可以作为输出设备,该输出设备被配置为分别向系统1000的用户输出视觉信息和听觉信息。

[0152] 供电设备1470可以适当地转换从嵌入在系统1000中的电池 (未示出) 和/或外部电源提供的电力,并将转换后的电力提供给系统1000的每个组件。

[0153] 连接接口1480可以提供系统1000与外部设备之间的连接,该外部设备连接到系统1000并且能够向系统1000发送数据以及从系统1000接收数据。连接接口1480可以通过使用各种接口方案来实现,诸如高级技术附件 (ATA)、串行ATA (SATA)、外部SATA (e-SATA)、小型计算机小型接口 (SCSI)、串行附加SCSI (SAS)、外围组件互连 (PCI)、PCI快速 (PCIe)、NVMe、IEEE 1394、通用串行总线 (USB) 接口、安全数字 (SD) 卡接口、多媒体卡 (MMC) 接口、eMMC接口、UFS接口、嵌入式UFS (eUFS) 接口和紧凑型闪存 (CF) 卡接口。

[0154] 根据本公开,可以提供一种被配置为基于电荷共享对多比特数据执行MAC运算的

SRAM单元。

[0155] 根据本公开,可以提供一种被配置为在不使用单独存储元件(比如电容器)的情况下执行MAC运算的SRAM单元。

[0156] 根据本公开,由于基于电荷共享对多比特数据执行MAC运算,可以获得良好的线性度。

[0157] 尽管已经参考本公开的实施例描述了本公开,但是对于本领域普通技术人员而言将显而易见的是,在不脱离所附权利要求所阐述的本公开的精神和范围的情况下,可以对其进行各种改变和修改。

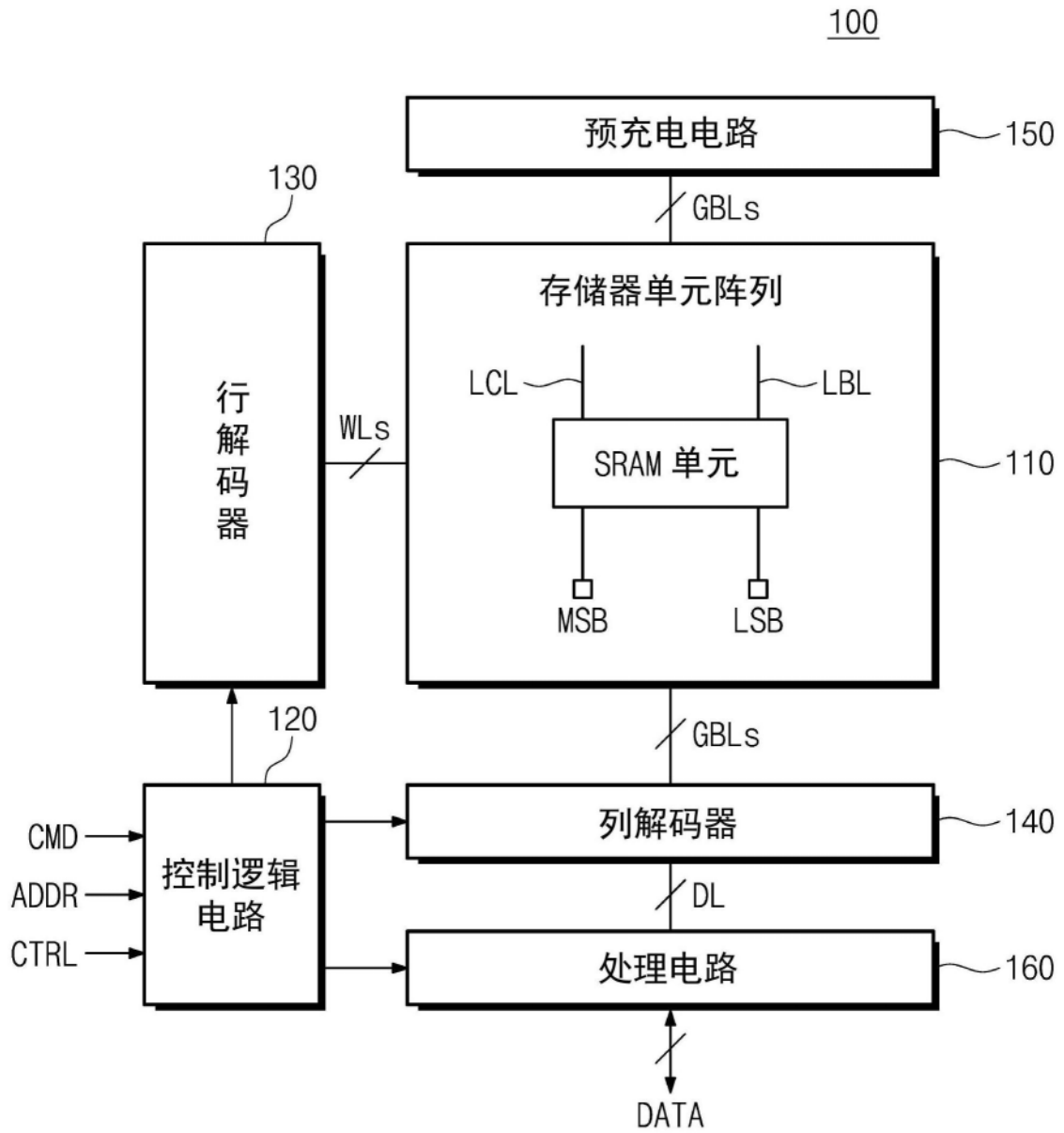


图1

情况	输入数据	WL 电平	权重	相乘结果		
				LCL	LBL	电荷共享后
-	iAct[1:0]	WLC/WLW	W			
1	任意值		0	放电到 CSS 或 VSS	VSS	VSS
2		0/0	1	放电到 VSS	VSS	VSS
3		0/1		放电到 VSS	VDD	1/3 x VDD
4		1/0		VDD	VSS	2/3 x VDD
5		1/1		VDD	VDD	VDD

图3B

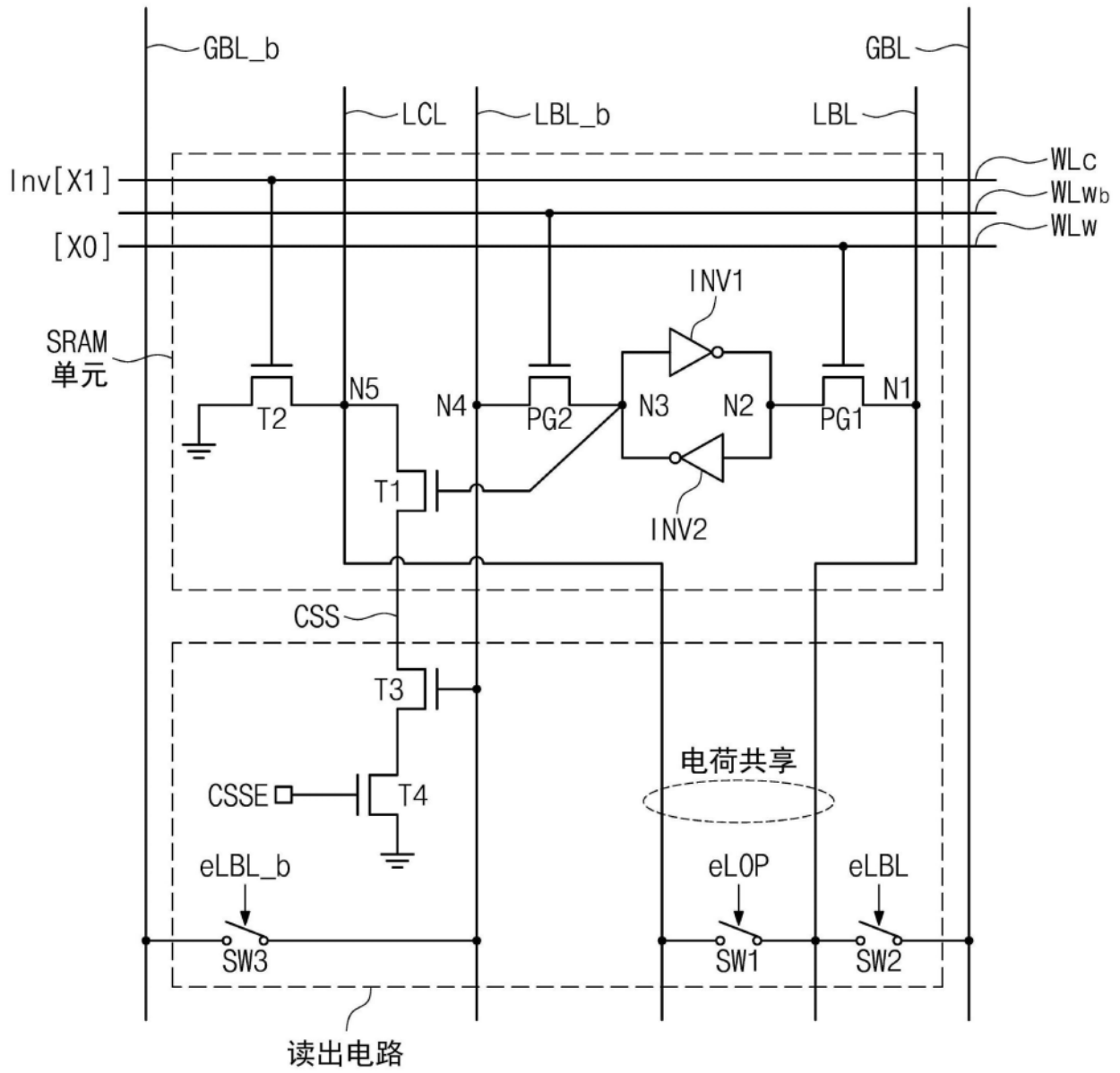


图4

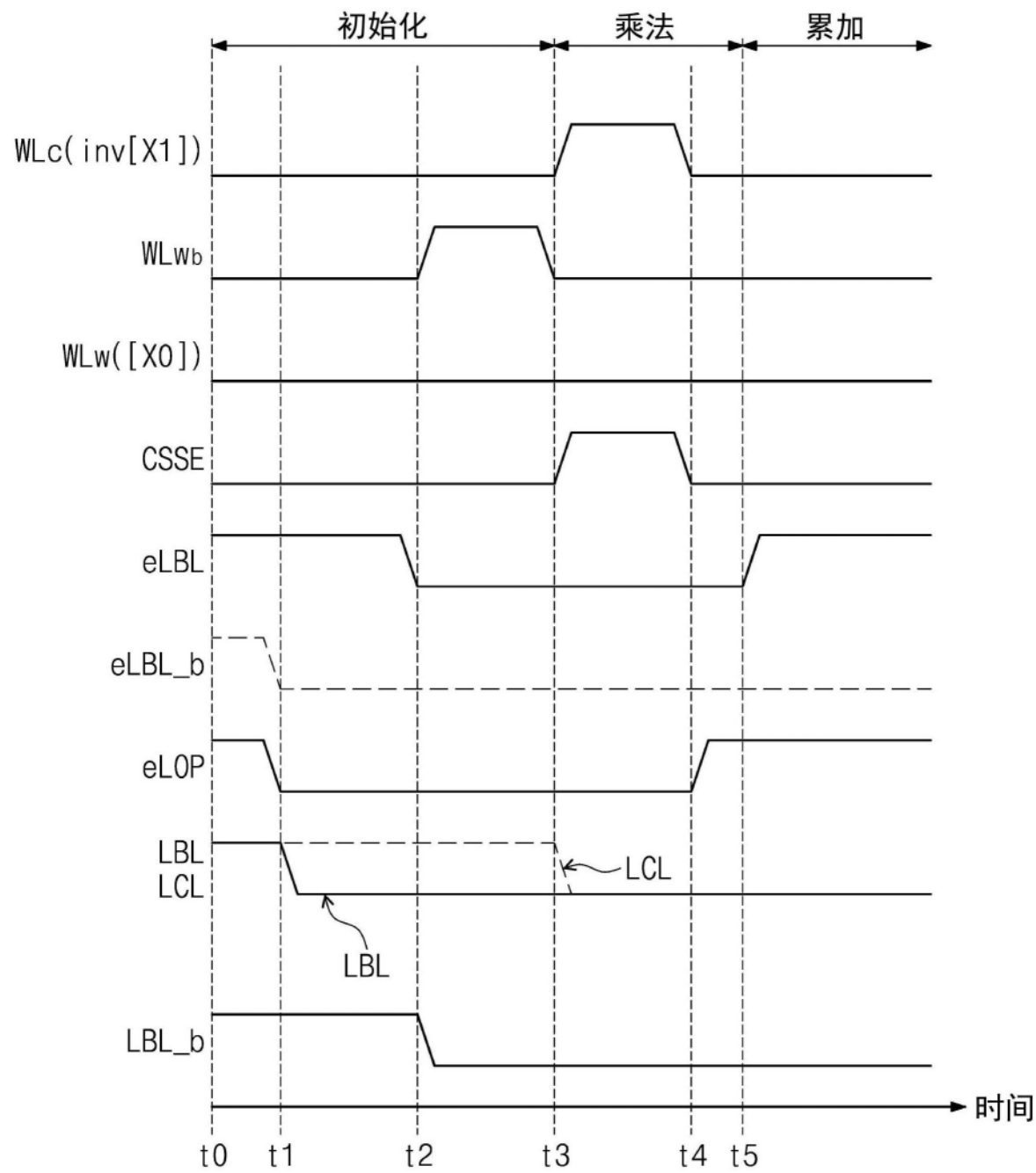


图5A

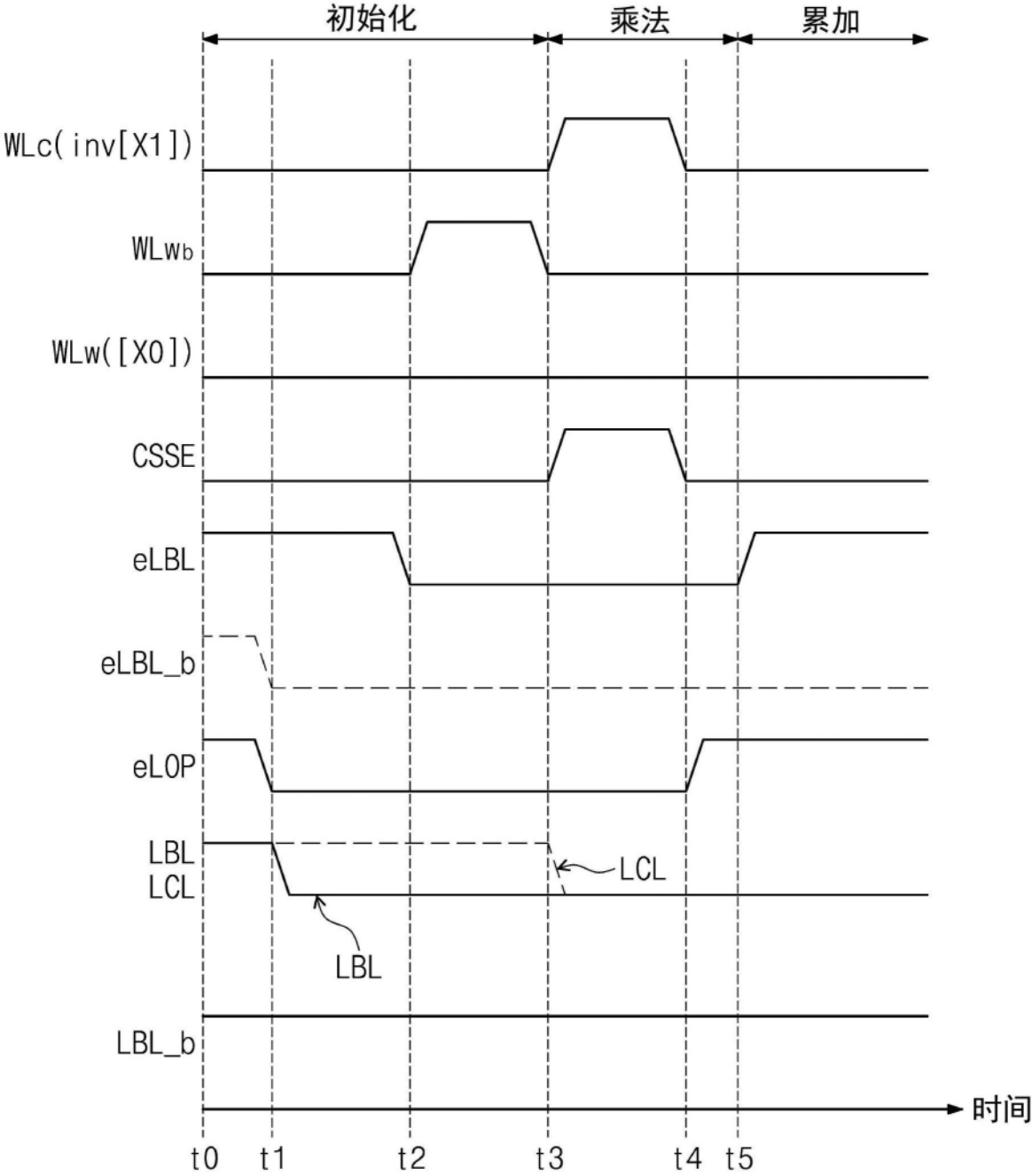


图5B

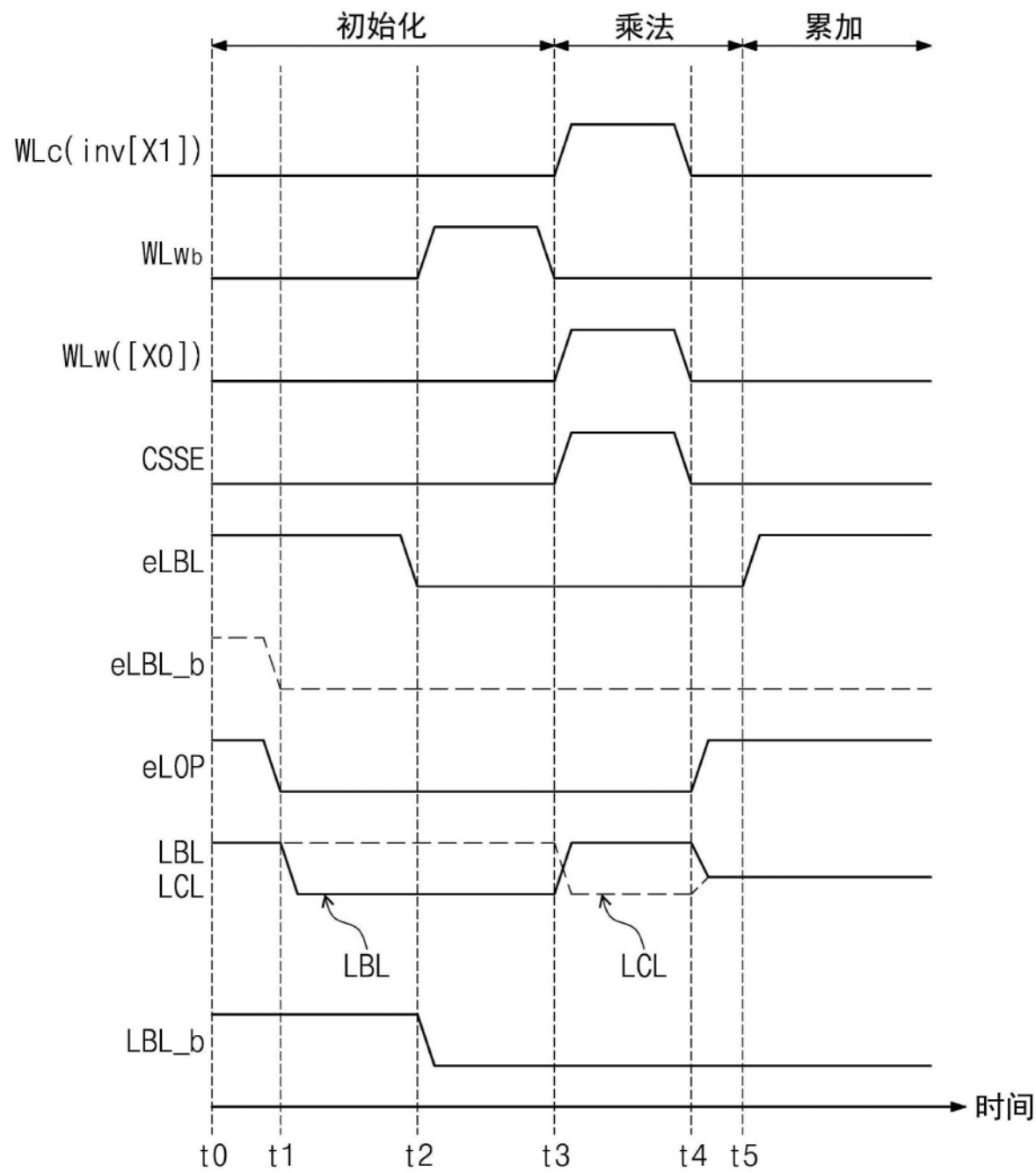


图6A

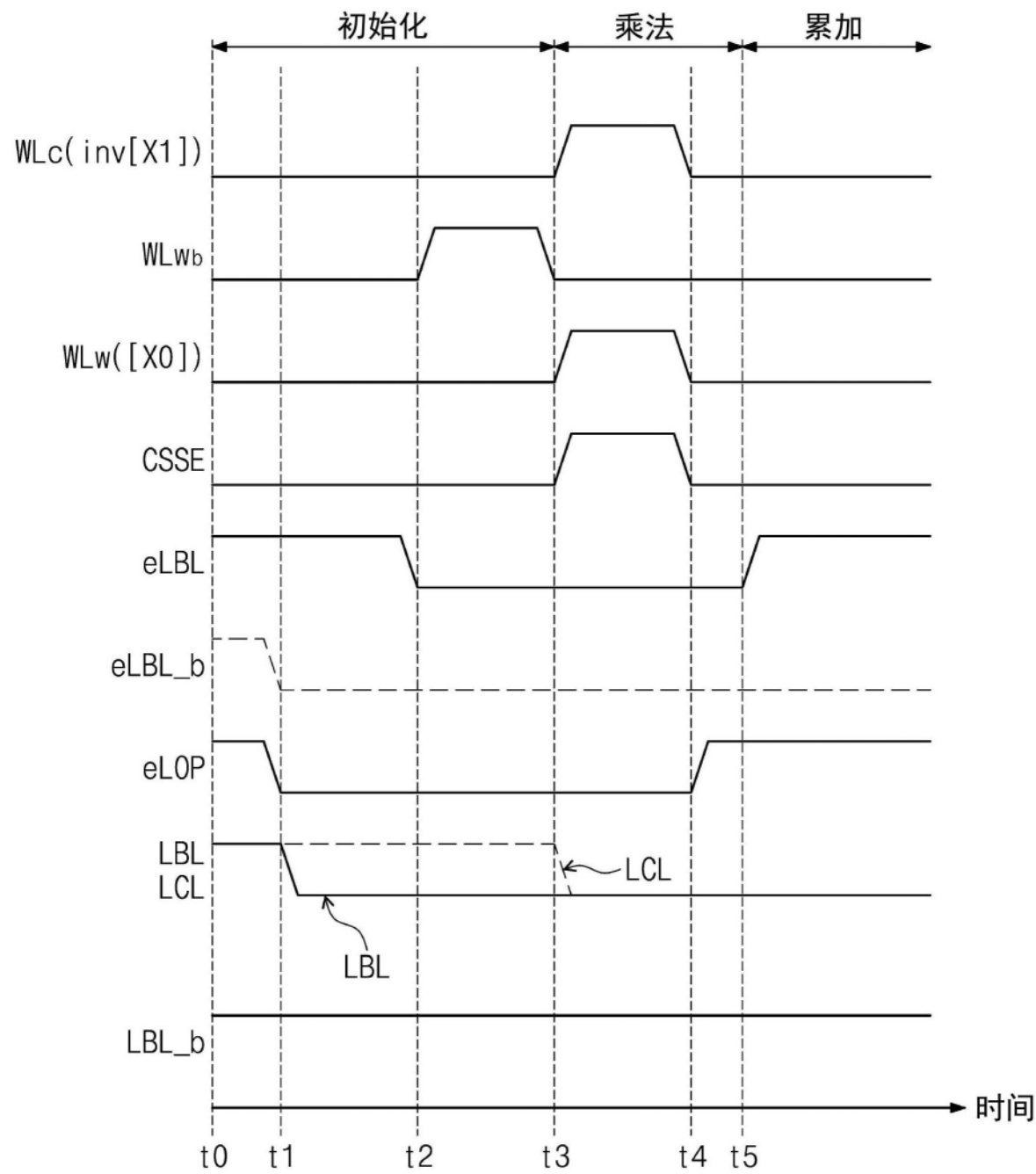


图6B

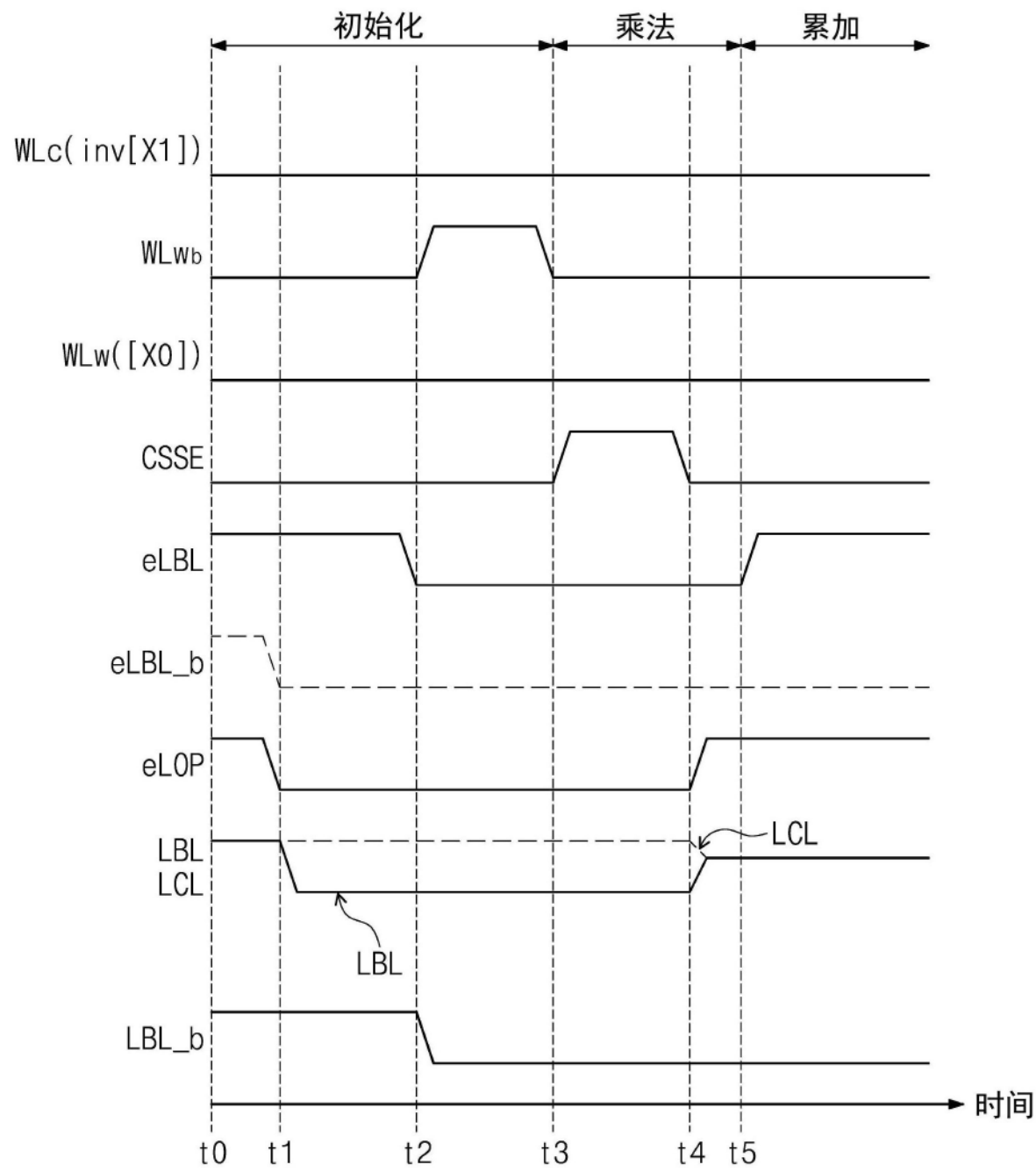


图7A

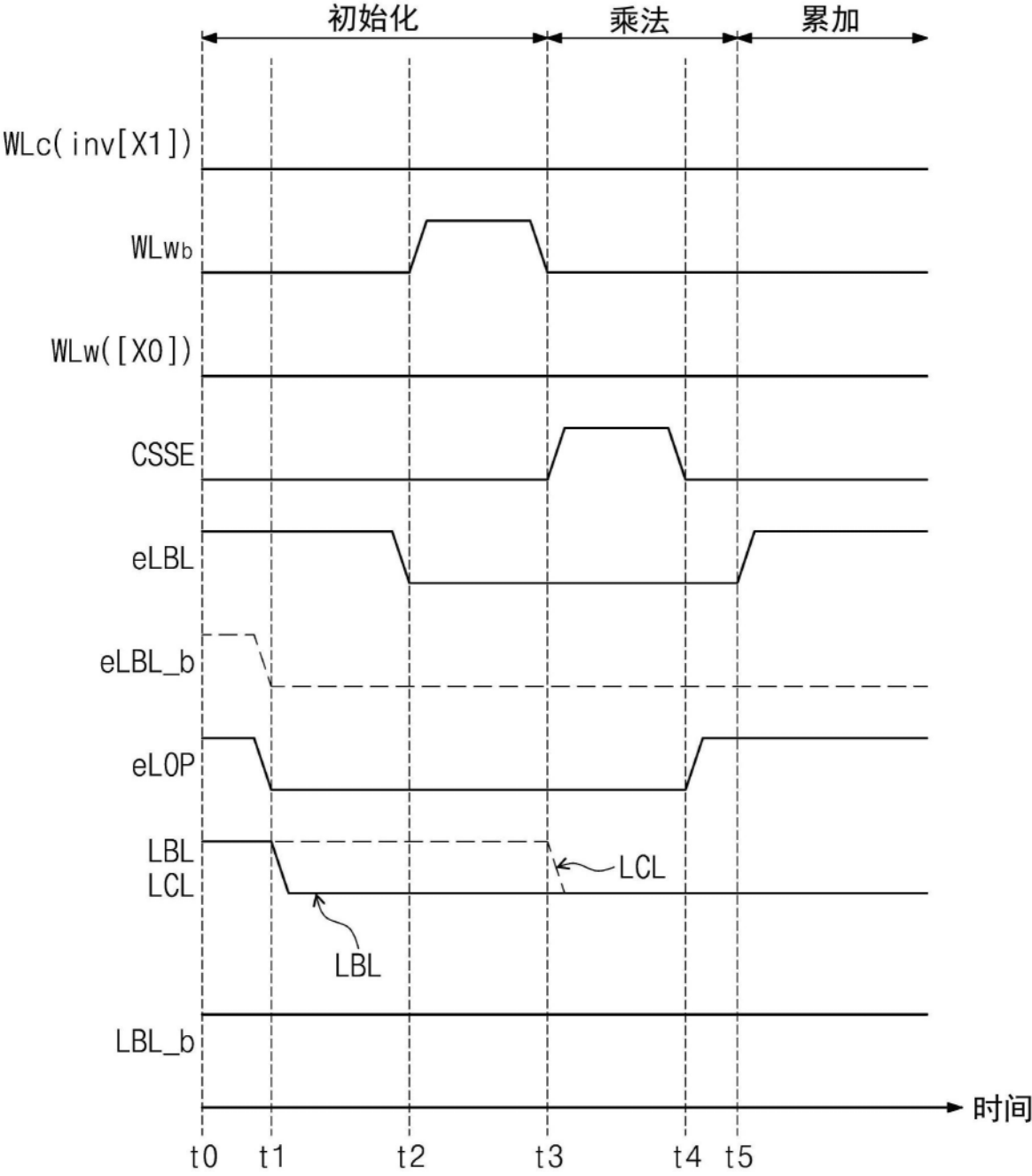


图7B

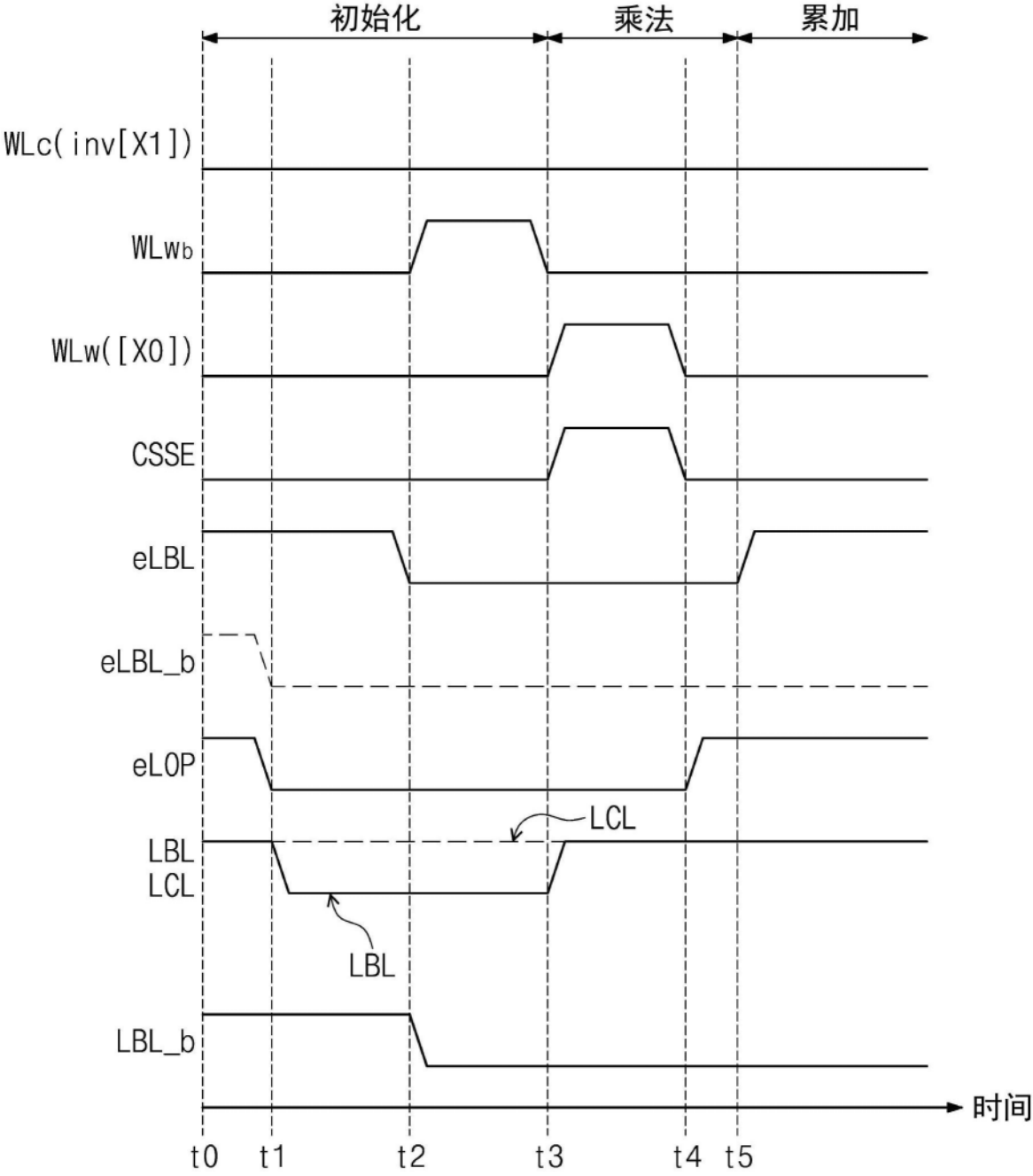


图8A

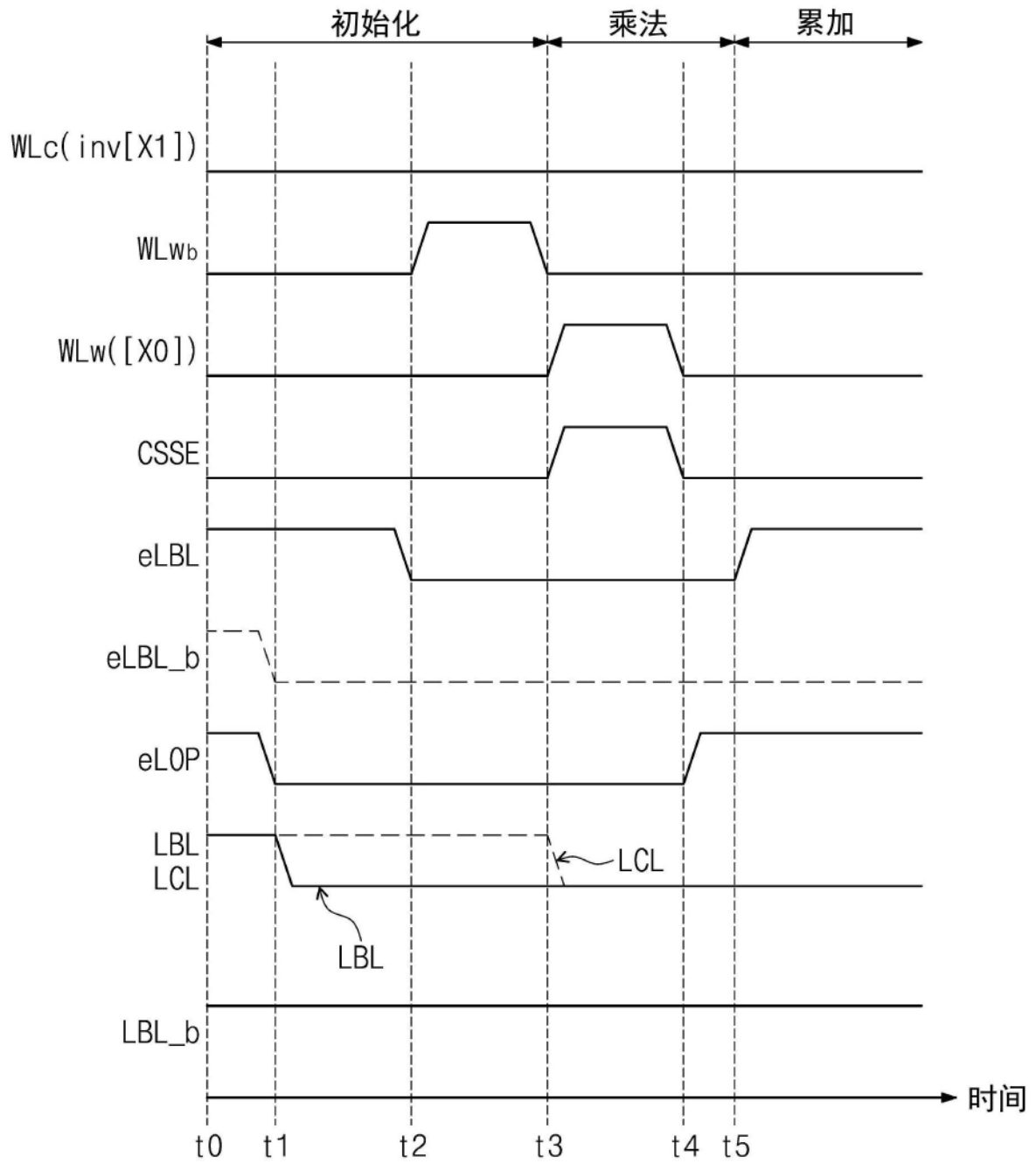


图8B

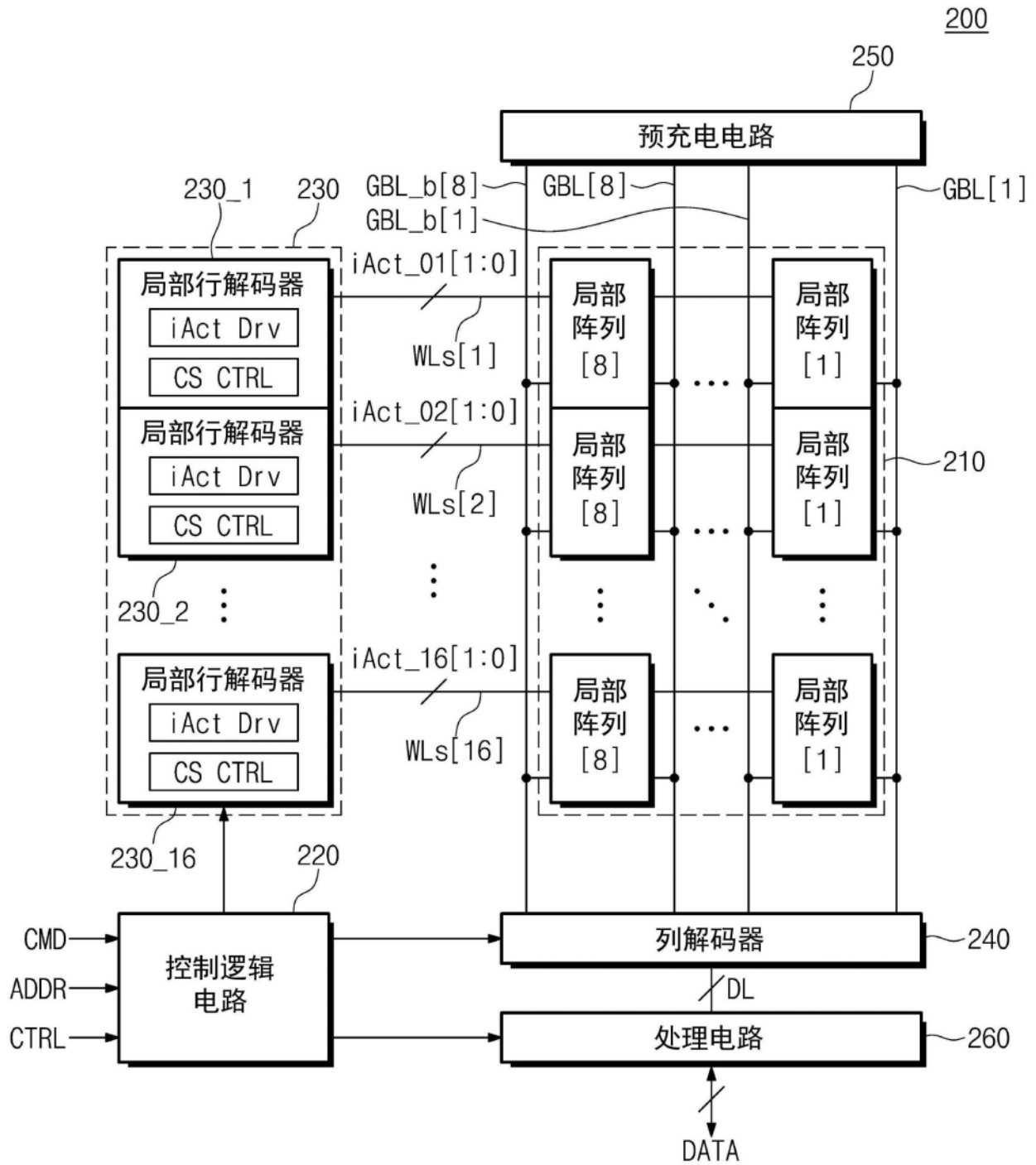


图9

局部阵列

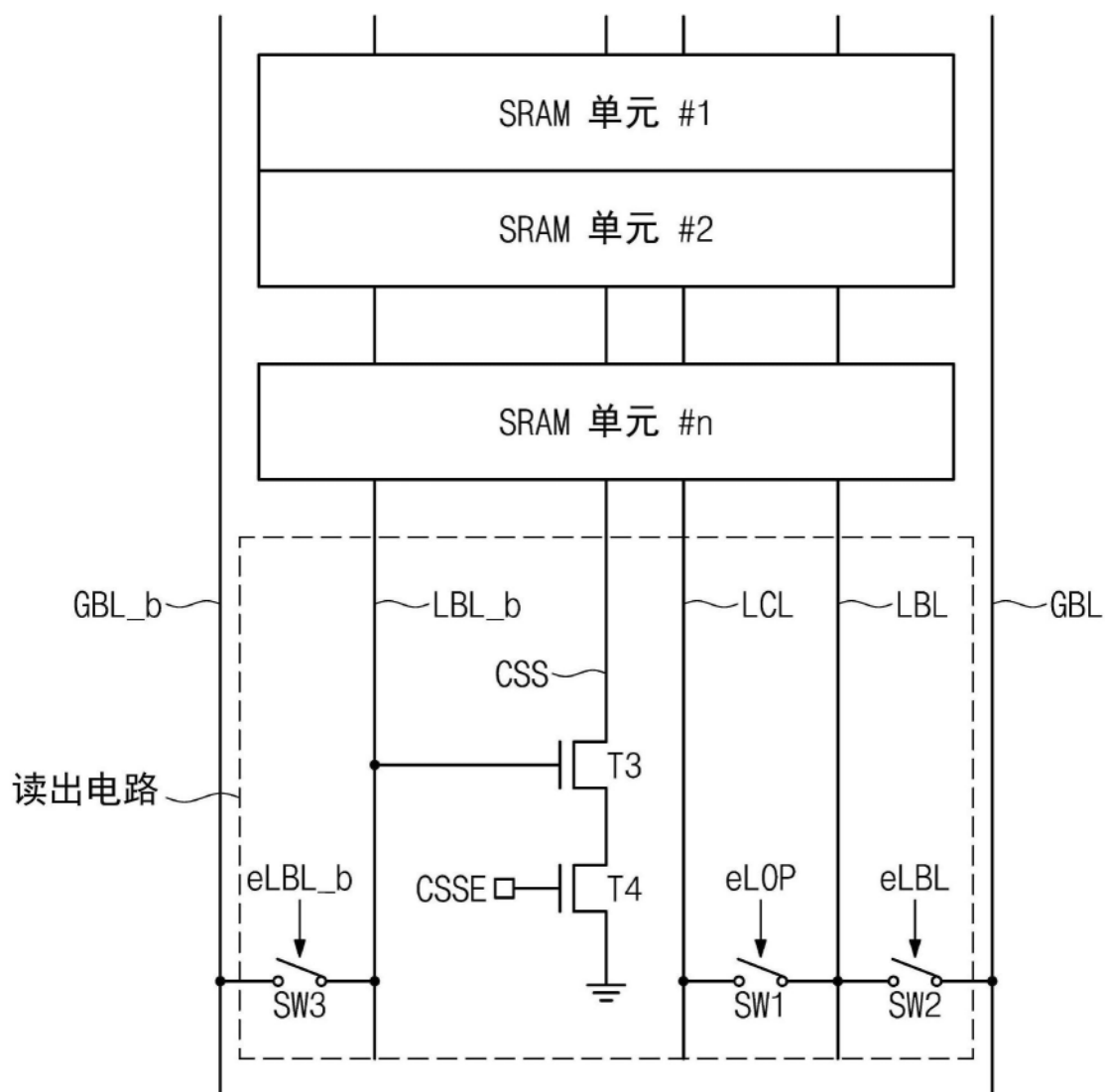


图10

200

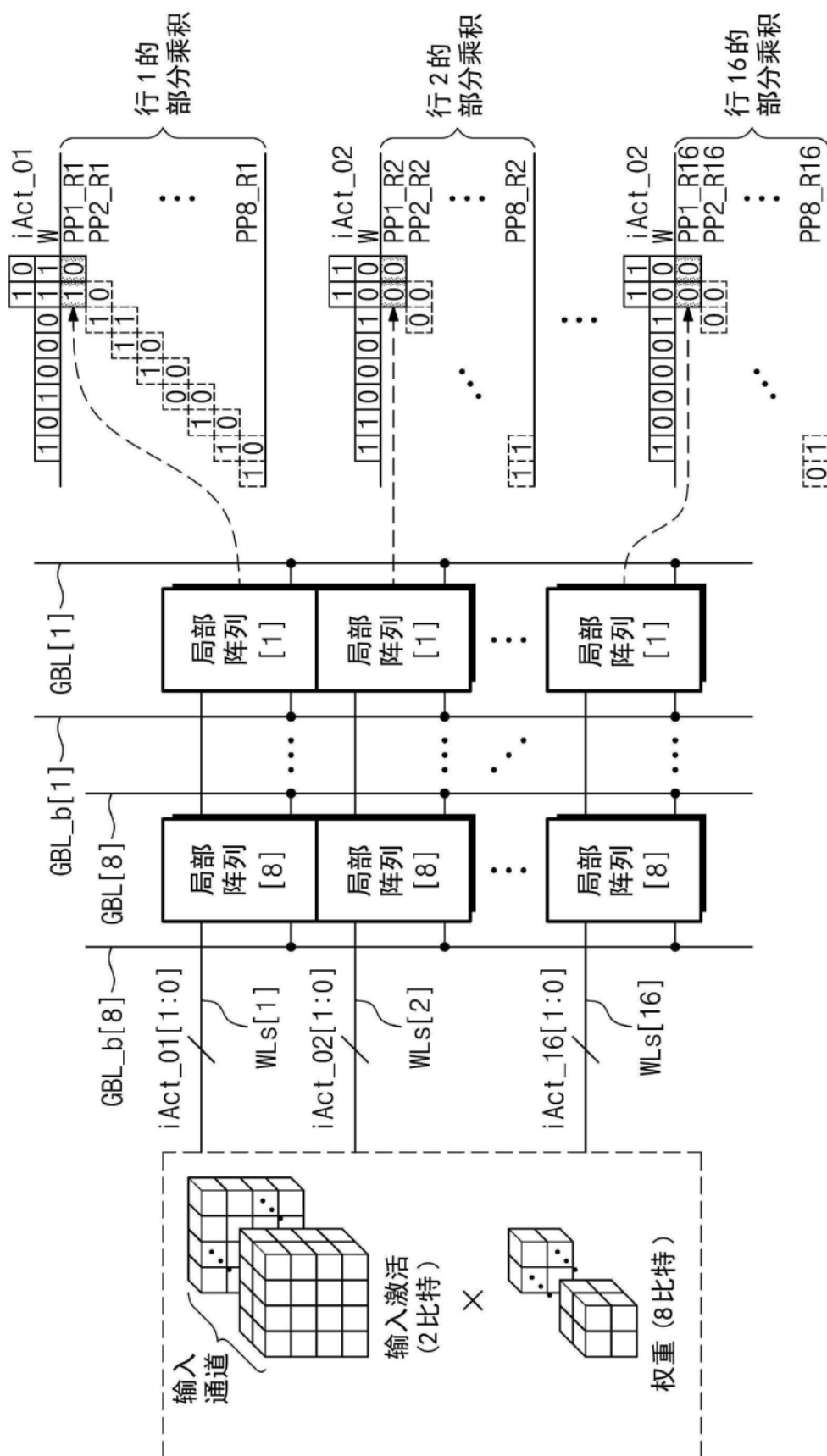


图11

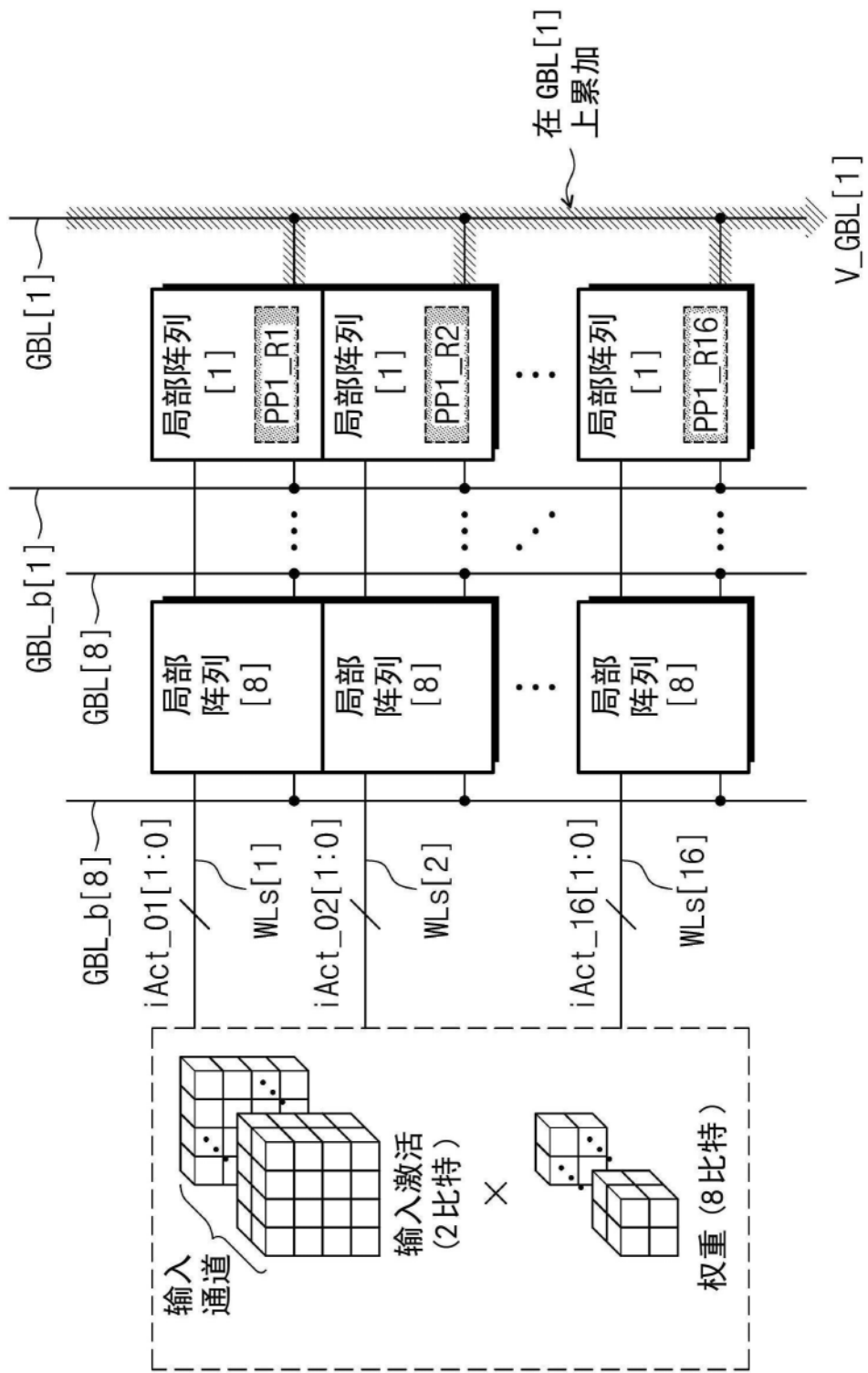


图12

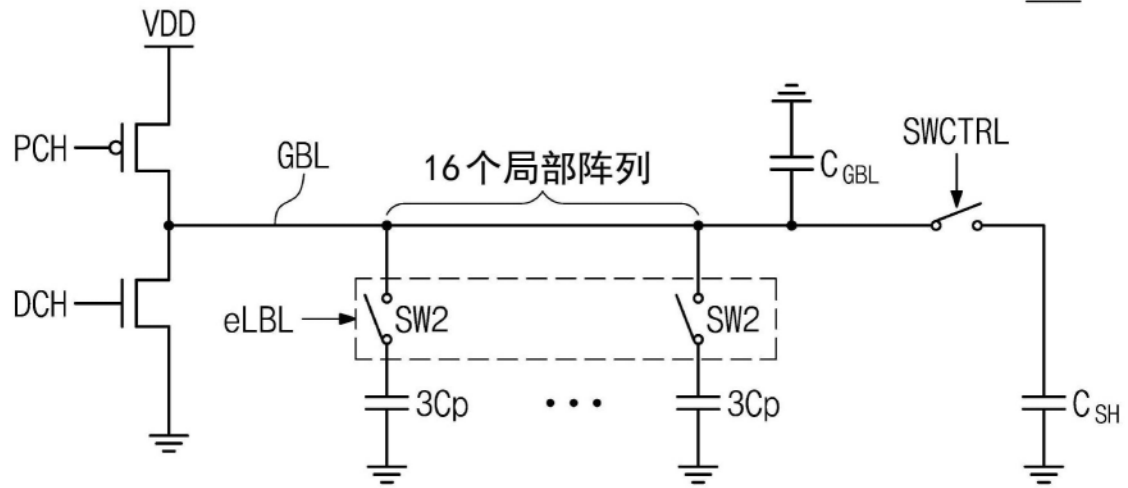
200

图13

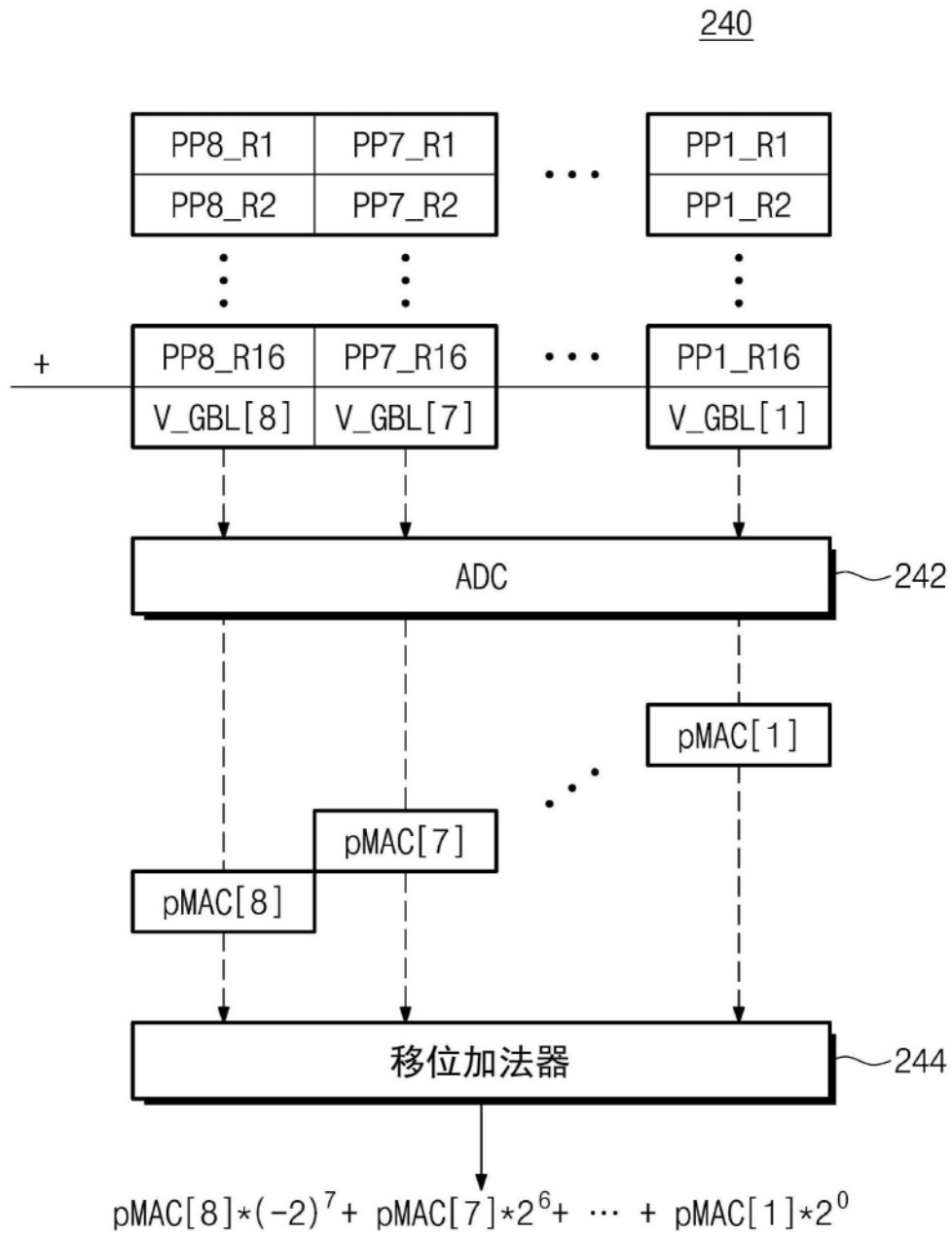


图14

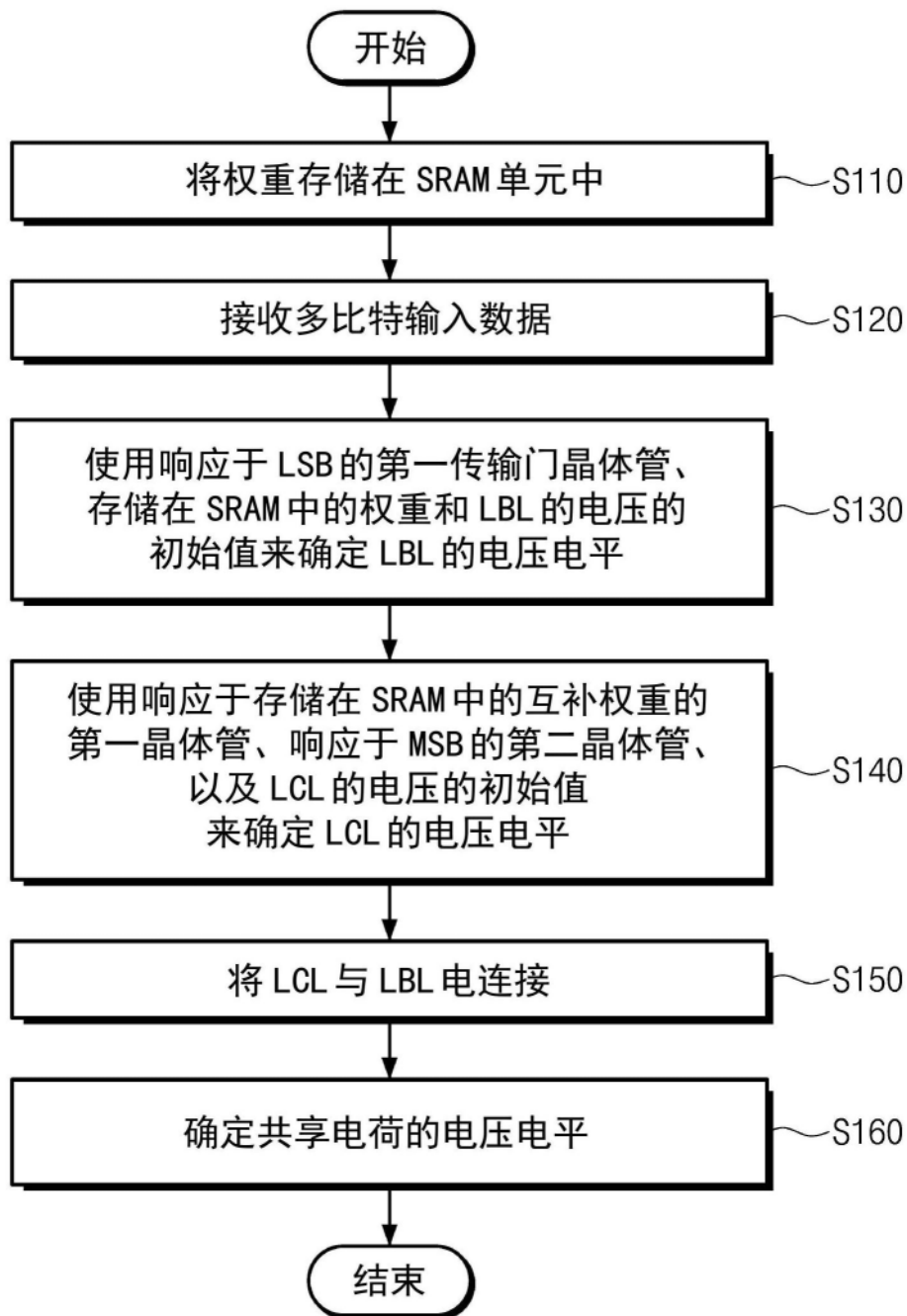


图15

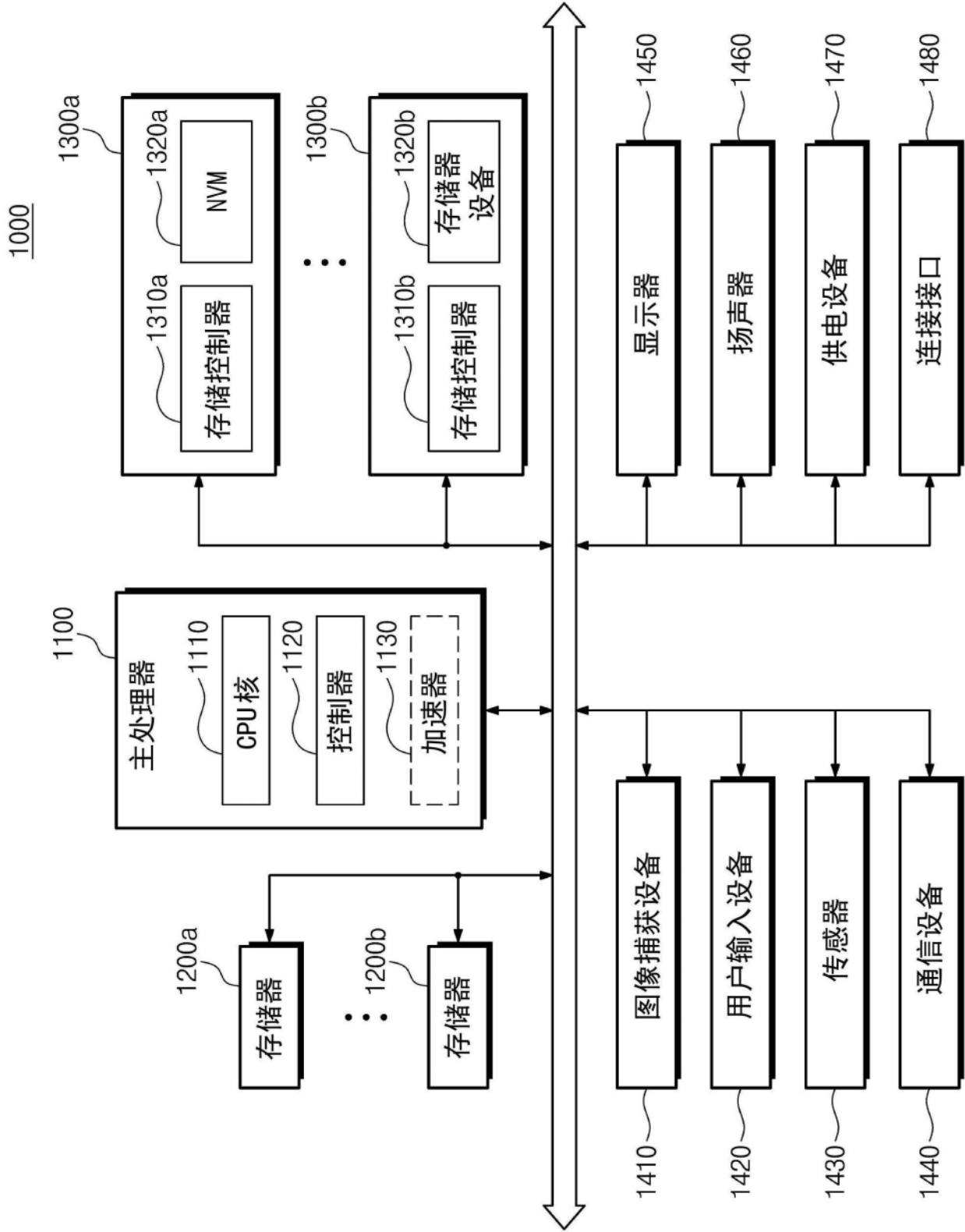


图16