



(12)发明专利申请

(10)申请公布号 CN 106847335 A

(43)申请公布日 2017.06.13

(21)申请号 201611235411.4

(22)申请日 2016.12.27

(71)申请人 北京大学

地址 100871 北京市海淀区颐和园路5号

(72)发明人 康晋锋 柳晨 黄鹏 周正

刘力锋 刘晓彦

(74)专利代理机构 中科专利商标代理有限责任

公司 11021

代理人 任岩

(51)Int.Cl.

G11C 13/00(2006.01)

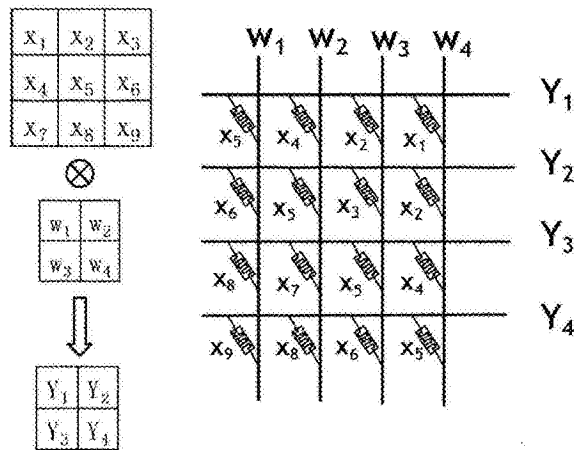
权利要求书2页 说明书8页 附图4页

(54)发明名称

基于阻变存储阵列的卷积计算存储一体化设备及方法

(57)摘要

本发明公开了一种并行的、存储计算一体化的卷积计算的硬件实现以及操作方式。该硬件实现主要基于阻变存储器二维交叉阵列,包括:阻变存储阵列、位线控制单元、字线控制单元、存储模块、输入模块和输出模块,存储模块产生输入矩阵信号,输入模块产生卷积核信号,位线控制单元用于选通存储模块或输入模块,字线控制单元用于选通输出模块或地线,以将输入矩阵存储至阻变存储阵列,以及将卷积核矩阵输入所述阻变存储阵列,输出模块输出卷积运算结果,以此方式实现卷积的存储计算一体化及并行计算。



1. 一种基于阻变存储阵列的卷积计算存储一体化设备,其特征在于,包括:  
阻变存储阵列、位线控制单元、字线控制单元、存储模块、输入模块和输出模块;  
所述存储模块产生输入矩阵信号;所述输入模块产生卷积核信号;  
所述位线控制单元用于选通存储模块或输入模块,字线控制单元用于选通输出模块或地线,以将输入矩阵存储至所述阻变存储阵列,以及将卷积核矩阵输入所述阻变存储阵列;  
所述输出模块输出卷积运算结果。
2. 根据权利要求1所述的卷积计算存储一体化设备,其特征在于,  
首先进入存储模式,位线控制单元选通存储模块,字线控制单元选通地线,在输入矩阵信号的作用下,输入矩阵被存储至所述阻变存储阵列;  
之后进入运算模式,位线控制单元选通输入模块,字线控制单元选通输出模块,与卷积核信号对应的卷积核矩阵输入所述阻变存储阵列,得到卷积运算结果。
3. 根据权利要求2所述的卷积计算存储一体化设备,其特征在于,  
所述阻变存储阵列包括交叉设置的多条位线和多条字线,以及多个阻变运算存储子单元,阻变运算存储子单元设置在位线和字线的交叉点处;  
所述位线控制单元和字线控制单元包括多个选通器。
4. 根据权利要求3所述的卷积计算存储一体化设备,其特征在于,  
所述存储模块包括多个信号发生器;  
在存储模式中,位线控制单元的选通器连通信号发生器与位线,字线控制单元的选通器接地;  
信号发生器产生输入矩阵信号,输入矩阵信号调整阻变运算存储子单元的电阻值,阻变运算存储子单元的电阻值代表输入矩阵的元素的大小,实现输入矩阵的存储。
5. 根据权利要求3所述的卷积计算存储一体化设备,其特征在于,  
所述输入模块包括多个信号发生器,  
在运算模式中,位线控制单元的选通器连通信号发生器与位线,字线控制单元的选通器选通输出模块;  
信号发生器产生卷积核信号,卷积核信号的电压值代表卷积核矩阵元素的大小,卷积核信号施加在阻变运算存储子单元上,进行卷积运算。
6. 根据权利要求2所述的卷积计算存储一体化设备,其特征在于,所述输入矩阵信号不变,改变所述卷积核信号,实现同一输入矩阵与不同卷积核矩阵的卷积运算。
7. 根据权利要求4或5所述的卷积计算存储一体化设备,其特征在于,当卷积核矩阵为 $k \times k$ 的矩阵,输入矩阵为 $m \times n$ 的矩阵时,作为卷积运算结果的输出矩阵为 $(m-k+1) \times (n-k+1)$ 的矩阵。
8. 根据权利要求7所述的卷积计算存储一体化设备,其特征在于,  
对于代表输出矩阵的元素 $Y(i, j)$ 的字线,依次选通第1条位线至第 $k \times k$ 条位线,存储模块信号发生器的电压依次加载到第1条位线至第 $k \times k$ 条位线,改变该条字线与第1条位线至第 $k \times k$ 条位线交叉处的阻变运算存储子单元的电阻值,使得该条字线与第1条位线至第 $k \times k$ 条位线交叉处的阻变运算存储子单元的电阻值分别代表输入矩阵中的元素,依次对各条字线进行上述操作,完成输入矩阵的存储。
9. 根据权利要求7所述的卷积计算存储一体化设备,其特征在于,

对于代表输出矩阵的元素 $Y(i, j)$ 的字线,同时选通第1条位线至第 $k \times k$ 条位线,输入模块信号发生器的电压同时加载到该条字线与第1条位线至第 $k \times k$ 条位线交叉处的阻变运算存储子单元,加载到该条字线与第1条位线至第 $k \times k$ 条位线交叉处的阻变运算存储子单元的电压值分别代表卷积核矩阵的元素,同时对各条字线进行上述操作,完成卷积核矩阵的输入。

10.一种卷积计算存储一体化方法,利用权利要求1-9任一项所述的卷积计算存储一体化设备,其特征在于,包括以下步骤:

位线控制单元选通存储模块,字线控制单元选通接地;

存储模块产生电压信号,将输入矩阵存储至阻变存储阵列,阻变运算存储子单元的电阻值代表所述输入矩阵的元素;

位线控制单元选通输入模块,字线控制单元选通输出模块;

输入模块产生电压信号,将卷积核矩阵输入阻变存储阵列,该电压信号值与卷积核矩阵元素值成正比;

将阻变存储阵列每条字线的阻变运算存储子单元的电流求和,并将求和后的输出电流转化为数字信号或电压信号

将数字信号或电压信号编码排列得到输出矩阵。

## 基于阻变存储阵列的卷积计算存储一体化设备及方法

### 技术领域

[0001] 本发明涉及半导体集成电路及其制造技术领域,更具体地,涉及一种基于阻变存储阵列的卷积计算存储一体化设备及方法。

### 背景技术

[0002] 卷积操作在图像处理中有着广泛的应用,例如图像去噪、特征提取、平滑、滤波、边缘检测、图像增强等。特别地,卷积操作在卷积神经网络(CNN)中具有关键作用,对输入图像的卷积操作是实现卷积神经网络特定功能的基础。作为目前应用十分广泛的人工神经网络算法,由于卷积神经网络具有并行性、分布式计算、自学习及自适应能力以及很强的鲁棒性和容错性等特点,在近些年得到了科学家们广泛的研究关注。传统的实现卷积的方法主要是基于现代计算机技术,即基于冯诺依曼体系架构和布尔代数逻辑的CMOS数字电路实现。在冯诺依曼体系构架中运算部分和存储部分是分离的,数据通过总线在运算部分和存储部分之间进行交换。数据在总线上进行交换会消耗能量和降低计算效率,随着基于CMOS技术的运算部分技术和存储部分技术的发展,数据在总线上的交换速度已经成为制约体系计算效率提高的瓶颈,同时数据通过总线交换过程的能耗造成体系的能耗增加。与此同时利用布尔逻辑进行卷积中的乘积求和计算时需要使用大量逻辑门电路,并且需要存储大量中间结果,这造成了高延时、高功耗以及额外的存储开销。在硬件消耗以及功耗的限制下,传统的卷积实现方法无法满足人们日益增长的低功耗、快速计算的需求。

### 发明内容

[0003] (一)要解决的技术问题

[0004] 为了突破现有的技术限制,提高卷积运算的速度及硬件利用效率,提出了一种基于阻变存储阵列实现卷积计算存储一体化的设备及方法。

[0005] (二)技术方案

[0006] 一种基于阻变存储阵列的卷积计算存储一体化设备,包括:

[0007] 阻变存储阵列、位线控制单元、字线控制单元、存储模块、输入模块和输出模块;

[0008] 所述存储模块产生输入矩阵信号;所述输入模块产生卷积核信号;

[0009] 所述位线控制单元用于选通存储模块或输入模块,字线控制单元用于选通输出模块或地线,以将输入矩阵存储至所述阻变存储阵列,以及将卷积核矩阵输入所述阻变存储阵列;所述输出模块输出卷积运算结果。

[0010] 上述方案中,首先进入存储模式,位线控制单元选通存储模块,字线控制单元选通地线,在输入矩阵信号的作用下,输入矩阵被存储至所述阻变存储阵列;

[0011] 之后进入运算模式,位线控制单元选通输入模块,字线控制单元选通输出模块,与卷积核信号对应的卷积核矩阵输入所述阻变存储阵列,得到卷积运算结果。

[0012] 上述方案中,所述阻变存储阵列包括交叉设置的多条位线和多条字线,以及多个阻变运算存储子单元,阻变运算存储子单元设置在位线和字线的交叉点处;

- [0013] 所述位线控制单元和字线控制单元包括多个选通器。
- [0014] 上述方案中,所述存储模块包括多个信号发生器;
- [0015] 在存储模式中,位线控制单元的选通器连通信号发生器与位线,字线控制单元的选通器接地;
- [0016] 信号发生器产生输入矩阵信号,输入矩阵信号调整阻变运算存储子单元的电阻值,阻变运算存储子单元的电阻值代表输入矩阵的元素的大小,实现输入矩阵的存储。
- [0017] 上述方案中,所述输入模块包括多个信号发生器,
- [0018] 在运算模式中,位线控制单元的选通器连通信号发生器与位线,字线控制单元的选通器选通输出模块;
- [0019] 信号发生器产生卷积核信号,卷积核信号的电压值代表卷积核矩阵元素的大小,卷积核信号施加在阻变运算存储子单元上,进行卷积运算。
- [0020] 上述方案中,所述输入矩阵信号不变,改变所述卷积核信号,实现同一输入矩阵与不同卷积核矩阵的卷积运算。
- [0021] 上述方案中,其特征在于,当卷积核矩阵为 $7 \times 7$ 的矩阵,输入矩阵为 $m \times n$ 的矩阵时,作为卷积运算结果的输出矩阵为 $(m-k+1) \times (n-k+1)$ 的矩阵。
- [0022] 上述方案中,对于代表输出矩阵的元素 $Y(i, j)$ 的字线,依次选通第1条位线至第七 $\times$ 七条位线,存储模块信号发生器的电压依次加载到第1条位线至第七 $\times$ 七条位线,改变该条字线与第1条位线至第七 $\times$ 七条位线交叉处的阻变运算存储子单元的电阻值,使得该条字线与第1条位线至第七 $\times$ 七条位线交叉处的阻变运算存储子单元的电阻值分别代表输入矩阵中的元素,依次对各条字线进行上述操作,完成输入矩阵的存储。
- [0023] 上述方案中,对于代表输出矩阵的元素 $Y(i, j)$ 的字线,同时选通第1条位线至第七 $\times$ 七条位线,输入模块信号发生器的电压同时加载到该条字线与第1条位线至第七 $\times$ 七条位线交叉处的阻变运算存储子单元,加载到该条字线与第1条位线至第七 $\times$ 七条位线交叉处的阻变运算存储子单元的电压值分别代表卷积核矩阵的元素,同时对各条字线进行上述操作,完成卷积核矩阵的输入。
- [0024] 一种卷积计算存储一体化方法,利用上述的卷积计算存储一体化设备,包括以下步骤:
- [0025] 位线控制单元选通存储模块,字线控制单元选通接地;
- [0026] 存储模块产生电压信号,将输入矩阵存储至阻变存储阵列,阻变运算存储子单元的电阻值代表所述输入矩阵的元素;
- [0027] 位线控制单元选通输入模块,字线控制单元选通输出模块;
- [0028] 输入模块产生电压信号,将卷积核矩阵输入阻变存储阵列,该电压信号值与卷积核矩阵元素值成正比;
- [0029] 将阻变存储阵列每条字线的阻变运算存储子单元的电流求和,并将求和后的输出电流转化为数字信号或电压信号
- [0030] 将数字信号或电压信号编码排列得到输出矩阵。
- [0031] (三)有益效果
- [0032] 利用阻变存储器件的阻值可调特性,该架构将电压信号作为卷积核信号、将阻变存储器的阻值作为输入矩阵信号,利用一定的编码方式将输入矩阵存入阻变存储器阵列,

将输出电流作为卷积结果,脱离了布尔逻辑框架,实现了计算存储一体化,降低功耗的同时大幅提升了运算速度,并且利用阻变存储器阵列可大规模集成的特点进行了并行扩展,实现了所有卷积结果同时并行输出。

### 附图说明

- [0033] 图1是本发明实施例的基于阻变存储阵列的卷积计算存储一体化设备的结构示意图;
- [0034] 图2是本发明实施例的阻变存储阵列示意图;
- [0035] 图3是本发明实施例的卷积计算存储一体化方法流程图;
- [0036] 图4是本发明实施例的卷积计算示意图。
- [0037] 图5是本发明实施例在存储模式下的电路连接示意图;
- [0038] 图6是本发明实施例在存储模式下的输入矩阵编码存储图;
- [0039] 图7是本发明实施例在运算模式下的电路连接示意图;
- [0040] 图8是本发明实施例的输入矩阵与卷积核矩阵的编码图。

### 具体实施方式

[0041] 本发明将待卷积运算的输入矩阵存入阻变存储阵列中,利用阻变存储阵列作为存储器件,将卷积核作为电压输入,实现了在存储器中直接进行卷积计算。

[0042] 为使本发明的目的、技术方案和优点更加清楚明白,以下结合具体实施例,并参照附图,对本发明作进一步的详细说明。

[0043] 以下,将参照附图来描述本公开的实例。但是应该理解,这些描述只是示例性的,而并非要限制本公开的范围。此外,在以下说明中,省略了对公知结构和技术的描述,以避免不必要地混淆本公开的概念。

[0044] 在附图中示出了根据本公开实例的各种结构示意图。这些图并非是按比例绘制的,其中为了清楚表达的目的,放大了某些细节,并且可能省略了某些细节。图中所示出的各种元件的形状以及它们之间的相对大小、位置关系仅是示例性的,实际中可能由于制造公差或技术限制而有所偏差,并且本领域技术人员根据实际所需可以另外设计具有不同形状、大小、相对位置的元件。

[0045] 实施例1

[0046] 本实施例提供一种基于阻变存储阵列的卷积计算存储一体化设备,利用阻变存储阵列通过电压/电流/电阻关系实现二维矩阵的存储、卷积计算一体化,以及卷积的并行运算。阻变存储阵列包括阻变器件,阻变器件的电阻值可以在两个不同的电阻值之间通过外加电压进行切换,称之为编程(program)/擦除(erase)(或者置位(SET)/复位(RESET))。编程电压和擦除电压极性相反,低于编程电压和擦除电压的电压信号不会改变阻变器件的电阻值。编程电压和擦除电压对应的两个电阻值的高低阻态有明显的区别,高阻态与低阻态的比值可达到 $10^4 \sim 10^5$ ,且低阻态在一定范围内可以连续取值。在本发明实施例中,输入电压作为卷积核、电阻值高低不同的电阻态作为计算和存储的待卷积计算的矩阵,输出电流(电压)作为计算结果。

[0047] 如图1所示给出了卷积计算存储一体化设备各个模块的组成以及各模块间的相互

关系,包括:阻变存储阵列、存储模块、输入模块、控制器、位线控制单元、字线控制单元以及输出模块。其相互连接、控制关系如下:

[0048] 如图2所示,阻变存储阵列包含沿第一方向延伸的多条字线、沿与第一方向交叉的第二方向延伸的多条位线以及设置于位线和字线的交叉点处的多个阻变运算存储子单元。

[0049] 存储模块包括多个信号发生器,信号发生器生成一个连续的正脉冲或负脉冲,产生输入矩阵信号,输入矩阵即为待卷积运算的矩阵。输入模块包括多个信号发生器,信号发生器生成一系列连续电压,产生卷积核信号。输出模块包括多个数模转换器或电流/电压转换器,输出正比于字线上输出电流的电压或电流信号,对输出信号进行处理,得到卷积运算后的输出矩阵。

[0050] 阻变存储阵列通过位线控制单元连接存储模块和输入模块,通过字线控制单元连接输出模块和接地。位线控制单元包含多个选通器,根据控制器给出的触发信号,通过选通器选通存储模块和输入模块,当选通存储模块时,存储模块的信号发生器与位线对应连接,当选通输入模块时,输入模块的信号发生器与位线对应连接,进行阻变存储子单元的阻态调整或者卷积运算。

[0051] 字线控制单元也包含多个选通器,根据控制器给出的触发信号,通过选通器选通接地或选通输出模块,当选通接地时,字线全部接地,当选通输出模块时,字线与数模转换器或电流/电压转换器对应连接。

[0052] 基于电阻运算和传递的原理,阻变存储阵列为如图2所示的交叉阵列。在交叉阵列中包含互相平行的多条位线、相互平行的多条字线,在位线和字线的每一交叉点处设置有一阻变运算存储子单元,阻变运算存储子单元一端连接位线,另一端连接字线;阻变运算存储子单元可以是单个阻变器件,也可以是单个阻变器件和双向选择器件串联的结构,例如如图2中的阻变器件和双向选择管串联结构,也可以是多个子单元并联的结构,每个子单元为阻变器件与三端开关串联的结构。字线和位线相互垂直,在交叉阵列同一行和同一列的阻变运算存储子单元可以直接构成运算单元,不同行不同列的子单元可以间接构成运算单元。

[0053] 阻变器件可以是基于氧化物的阻变存储器RRAM,基于金属离子氧化还原的阻变存储器CBRAM、相变存储器PCM、磁阻存储器MRAM等依靠电阻变化进行信息存储的元器件。

[0054] 当卷积计算存储一体化设备处于存储模式时,控制器发出相应的触发信号,位线控制单元选通存储模块,字线控制单元选通接地,存储模块的信号发生器与位线一一对应连接,存储模块产生的脉冲信号对阻变存储子单元的电阻态进行训练,利用不同脉冲个数调整阻变存储子单元的电阻态,即电导值,使其正比于待计算的输入矩阵中元素的值,即实现输入矩阵的存储。当处于运算模式时,位线控制单元选通输入模块,字线控制单元选通输出模块,输入模块产生与待进行卷积运算的卷积核成正比的电压信号,输出模块输出的电压或电流信号即是卷积计算的结果,实现卷积计算。

[0055] 将字线数目扩展为输出矩阵元素个数,每条字线上的阻变存储子单元按一定编码方式逐条字线将输入矩阵进行存储,在运算模式下施加输入电压后即可同时得到所有输出的卷积结果,以此方式实现了卷积的并行计算以及存储计算一体化。

[0056] 对于大小为 $m \times n$ 的输入矩阵以及大小为 $k \times k$ 的卷积核,输出卷积矩阵大小为 $(m-k+1) \times (n-k+1)$ ,所需的阻变存储阵列大小为 $k \times k \times (m-k+1) \times (n-k+1)$ ,所述输入矩阵及相

应卷积核矩阵的编码方式为:每条字线的输出代表输出矩阵中的一个元素 $Y(i, j)$ ,在元素 $Y(i, j)$ 所对应的字线上需要存储 $7 \times 7$ 大小的部分输入矩阵,将该部分输入矩阵一维展开为: $X(i+k-1, j+k-1), X(i+k-1, j+k-2), \dots, X(i+k-1, j), X(i+k-2, j+k-1), \dots, X(i, j+1), X(i, j)$ ,按顺序存储到位于该条字线上的阻变运算存储子单元中;将卷积核矩阵一维展开: $W(1,1), W(1,2), \dots, W(1,k), W(2,1), \dots, W(k, k-1), W(k, k)$ ,按顺序将代表卷积核矩阵的电压施加在位线上。

[0057] 本发明将待卷积运算的矩阵存入阻变存储器阵列中,利用阻变存储器阵列作为存储器件,将卷积核作为电压输入,如果卷积核矩阵中的元素的值为负时,直接输入负电压。

[0058] 本发明对同一输入矩阵做不同卷积运算时,若卷积核大小相同,只需改变输入电压,由于改变输入电压较为容易,重写阻变存储阵列存储的内容操作较为繁复,所以本发明更适用于对已经存储好的矩阵进行不同卷积计算的操作,利用对同一幅数字图片进行滤波、边缘检测等不同卷积处理。

[0059] 实施例2

[0060] 本发明提供一种卷积计算存储一体化方法,如图3所示,包括以下步骤:

[0061] S1:位线控制单元选通存储模块,字线控制单元选通接地。

[0062] S2:存储模块产生电压信号,将输入矩阵存储至阻变存储阵列,阻变运算存储子单元的电阻值代表所述输入矩阵的元素。

[0063] S3:位线控制单元选通输入模块,字线控制单元选通输出模块。

[0064] S4:输入模块产生电压信号,将卷积核矩阵输入阻变存储阵列,该电压信号值与卷积核矩阵元素值成正比。

[0065] S5:将阻变存储阵列每条字线的阻变运算存储子单元的电流求和,并将求和后的输出电流转化为数字信号或电压信号。

[0066] S6:将数字信号或电压信号编码排列得到输出矩阵。

[0067] 附图4给出了本发明适用的卷积计算种类以及计算方法。卷积是一个广义的积分概念,在数字信号处理、概率论、图像处理等领域都有广泛的应用。本发明只适用于二维离散卷积的计算,主要应用于二维图像的处理以及卷积神经网络(CNN)。图4以一个 $5 \times 5$ 的输入矩阵、 $3 \times 3$ 的卷积核为例进行二维离散卷积计算的步骤说明,本实例只作为举例,实际应用中输入及卷积核矩阵可扩展成任意的 $m \times n$ 矩阵,前提是输入矩阵面积大于卷积核矩阵。在图4中,输入矩阵中的元素为 $x_1 \sim x_{25}$ ,卷积核元素为 $w_1 \sim w_9$ ,首先将卷积核矩阵旋转 $180^\circ$ ,然后将翻转后的卷积核首先与输入矩阵的左上角部分重叠,之后计算该重叠部分的对应元素乘积并求和,作为第一个输出结果,即输出矩阵中的元素 $Y_1$ ;之后翻转后的卷积核向x方向平移一位,继续计算重叠部分的乘积和作为下一个输出结果,直到x方向全部重叠过,则卷积核向y方向平移一位并平移到x方向的起始点进行计算,重复以上过程直至输入矩阵中的元素全被卷积核覆盖过。在本例中,为防止卷积核溢出,不计算卷积核不能完全与输入矩阵重叠时的边缘结果,输入矩阵为 $5 \times 5$ ,卷积核为 $3 \times 3$ ,即输出矩阵为一 $3 \times 3$ 的矩阵。

[0068] 本发明实施例的卷积计算存储一体化方法具体如下:

[0069] S1:位线控制单元选通存储模块,字线控制单元选通接地。

[0070] 首先控制器给出触发信号控制位线控制单元选通存储模块、控制字线控制单元选通接地,如附图5所示。



[0071] S2:存储模块产生电压信号,将输入矩阵存储至阻变存储阵列,阻变运算存储子单元的电阻值代表所述输入矩阵的元素。

[0072] 对每个阻变运算存储子单元的电阻值进行调整,来储存输入矩阵。存储时每次选通一条字线及一条位线,使选通的字线连接其对应的存储单元的信号发生器,位线接地,利用存储模块信号发生器产生的电压调整被选通的字线与位线交叉点处的阻变运算存储子单元的电阻值。存储模块的信号发生器产生足够大的直流电压或多个电压脉冲,具体电压大小视应用的阻变存储单元的材料及待存储的输入矩阵元素值而定。通过不断的擦除、编程过程对阻变运算存储子单元的电阻值进行调整直至符合设定值。在计算电阻设定值时,应首先测得阻变器件电阻的变化范围,尤其是其最低电阻值。令该最低电阻值与输入矩阵中的最大元素相对应,求出比值,再进行输入矩阵其他元素与电阻值的反比对应。

[0073] 为了高效地计算卷积结果,要按照一定的编码方式存储输入矩阵。具体编码方式如附图6所示。此编码方式与具体使用的卷积核相关,对于给定的七×七大小的卷积核W,大小为 $m \times n$ 的输入矩阵X,输出矩阵Y大小应为 $(m-k+1) \times (n-k+1)$ ,所需的阻变存储阵列大小为 $k \times k \times (m-k+1) \times (n-k+1)$ 。

[0074] 每条字线的输出代表输出矩阵中的一个元素 $Y(i, j)$ ,在元素 $Y(i, j)$ 所对应的字线上需要存储七×七大小的部分输入矩阵,将该部分输入矩阵一维展开为: $X(i+k-1, j+k-1), X(i+k-1, j+k-2), \dots, X(i+k-1, j), X(i+k-2, j+k-1), \dots, X(i, j+1), X(i, j)$ ,按照以下编码存储输入矩阵:首先选通一条字线,该条字线的输出值代表输出矩阵中的元素 $Y(i, j)$ ,然后依次选通第1条位线至第七×七条位线,存储模块信号发生器的电压依次加载到第1条位线至第七×七条位线,改变该条字线与第1条位线至第七×七条位线交叉处的阻变运算存储子单元的电阻值,使得该条字线与第1条位线至第七×七条位线交叉处的阻变运算存储子单元的电阻值分别代表输入矩阵中的元素 $X(i+k-1, j+k-1), X(i+k-1, j+k-2), \dots, X(i+k-1, j), X(i+k-2, j+k-1), \dots, X(i, j+1), X(i, j)$ ,依次对各条字线进行上述操作,完成输入矩阵的存储。

[0075] S3:位线控制单元选通输入模块,字线控制单元选通输出模块。

[0076] 将每个阻变存储单元的电阻值调整之后,即将输入矩阵存入阻变存储阵列后,即可进行运算,即控制器给出触发信号控制位线控制单元选通输入模块、控制字线控制单元选通输出模块。

[0077] S4:输入模块产生电压信号,将卷积核矩阵输入阻变存储阵列,该电压信号值与卷积核矩阵元素值成正比。

[0078] 如附图7所示,控制器控制输入模块的信号发生器产生电压信号,电压信号值与卷积核矩阵元素值成正比关系,由于输出矩阵中的每个元素为卷积核矩阵元素与输入矩阵中的对应元素相乘然后相加的结果,所以将七×七大小的卷积核矩阵的元素进行一维展开为: $W(1,1), W(1,2), \dots, W(1,k), W(2,1), \dots, W(k,k-1), W(k,k)$ 。

[0079] 按照以下方式输入卷积核矩阵:同时选通所有字线与输出模块相连,字线上的输出值代表输出矩阵中的元素 $Y(i, j)$ ,然后同时选通第1条位线至第七×七条位线,输入模块信号发生器的电压同时加载到该条字线与第1条位线至第七×七条位线交叉处的阻变运算存储子单元,加载到该条字线与第1条位线至第七×七条位线交叉处的阻变运算存储子单元的电阻值分别代表卷积核矩阵的元素 $W(1,1), W(1,2), \dots, W(1,k), W(2,1), \dots, W(k,k-$

1),  $W(k, k)$ , 完成卷积核矩阵的输入。

[0080] 需要注意的是输入模块信号发生器的最大电压信号不能大于阻变运算存储子单元的编程/擦写电压, 以防止运算过程改变已存入的输入矩阵。

[0081] S5: 将每条字线的阻变运算存储子单元的电流相加求和, 并将相加求和后的输出电流转化为数字信号或电压信号。

[0082] 每条字线上的阻变运算存储子单元所存储的输入矩阵元素与位线上输入的卷积核矩阵元素一一对应, 加载到各条位线的电压与该条字线上的阻变运算存储子单元作用得到电流, 该条字线的输出电流即为各个阻变运算存储子单元的电流和, 该电流和即为该条字线代表的输出矩阵中的元素  $Y(i, j)$ 。

[0083] 如附图8所示, 以  $3 \times 3$  的输入矩阵、 $2 \times 2$  的卷积核矩阵为例说明, 输入矩阵的元素为  $x_1 \sim x_9$ , 卷积核元素为  $w_1 \sim w_4$ , 输入模块产生的电压信号与卷积核元素成正比, 即  $V_i = W_i$ ; 输入矩阵按上述编码方式存储至阻变存储阵列, 输入矩阵的元素值则与相应位置的阻变运算存储子单元的电阻成反比, 即与其电导成正比, 即  $R_j = 1/x_j$ 。根据欧姆定律有:  $I_i = V_i/R_j = x_j w_i$ , 即阻变运算存储子单元上的电流正比于输入矩阵中元素与卷积核中对应的元素的乘积; 利用电流的基尔霍夫定律, 同一字线上流过的电流即为乘积和, 即卷积。这样本发明利用硬件的物理特性直接实现了乘积和的计算, 即卷积计算, 并且实现了在存储矩阵的阵列中直接计算, 即存储计算一体化, 相比于基于 CMOS 电路的软件计算, 极大地减小了功耗及器件消耗。

[0084] 图7给出了一个优选的输出模块功能示意图, 输出模块包括多个模数转换器, 字线的输出电流为模拟电流, 模数转换器用于将模拟电流转换为数字电流。输出模块按需求也可以是一个电流/电压转换器件, 用于将输出电流成比例转换为输出电压, 用于下一级计算或存储设备。

[0085] S6: 将数字信号或电压信号编码排列得到输出矩阵。

[0086] 在这种架构下卷积的并行方式如附图8所示。由于在卷积运算中, 卷积核进行平移, 输入矩阵中的元素会多次被卷积核所覆盖并与卷积核中的不同位置的元素相乘。传统方法是利用时序信号, 将待运算矩阵分若干次输入, 每次输入矩阵的一部分, 是一种串行计算的方法。在本发明中, 利用将输入矩阵按一定编码方式进行重复存储的方法, 使得每条字线上输出一个卷积结果, 完成卷积的并行计算。具体举例来说, 在附图8中, 第1条字线输出电流得到第一个卷积输出  $Y_1$ , 从附图8可知  $Y_1 = x_1 w_4 + x_2 w_3 + x_4 w_2 + x_5 w_1$ , 即矩阵中只有  $x_1, x_2, x_4, x_5$  在此卷积计算中起作用; 第二条字线上输出电流得到第二个卷积输出  $Y_2$ ,  $Y_2 = x_6 w_4 + x_5 w_3 + x_3 w_2 + x_2 w_1$ , 即矩阵中有  $x_2, x_3, x_5, x_6$  在此字线上起作用。以此类推, 可利用4条字线完成示例中的卷积同时得到输出。附图8中的实例只作为举例, 实际应用中矩阵及卷积核可扩展成任意的  $m \times n$  矩阵, 前提是输入矩阵规模大于卷积核矩阵。利用这种方式, 在阻变存储阵列中实现了计算, 同时实现了所有卷积结果的输出, 即实现了卷积的并行计算, 极大地加快了运算速度。这种方式对阻变存储阵列大小的要求是: 位线数目不小于卷积核的元素个数, 字线数目不小于输出矩阵元素个数。

[0087] 在根据本发明的上述阻变运算存储设备中, 利用阻变存储器存储矩阵并进行卷积计算, 由于阻变存储器的可擦写性及非挥发性, 不需要电压维持, 并且支持矩阵的重写, 同时可利用改变输入电压的方式改变卷积核, 提高了器件利用率。

[0088] 本发明将输入矩阵存入阻变存储阵列中,利用阻变存储阵列作为存储器件,将卷积核矩阵作为输入,实现了在存储器件中直接计算。

[0089] 本发明对同一输入矩阵做不同卷积运算时,若卷积核大小相同,只需改变输入电压。由于改变输入电压较为容易,重写阻变存储阵列的存储内容操作繁复,所以本发明更适用于对已经存好的输入矩阵进行不同卷积计算的操作,例如对同一幅数字图片进行滤波、边缘检测等不同卷积处理。

[0090] 尽管已经参考本发明的典型实施例,具体示出和描述了本发明,但本领域普通技术人员应当理解,在不脱离所附权利要求所限定的本发明的精神和范围的情况下,可以对这些实施例进行形式和细节上的多种改变。

[0091] 以上所述的具体实施例,对本发明的目的、技术方案和有益效果进行了进一步详细说明,应理解的是,以上所述仅为本发明的具体实施例而已,并不用于限制本发明,凡在本发明的精神和原则之内,所做的任何修改、等同替换、改进等,均应包含在本发明的保护范围之内。

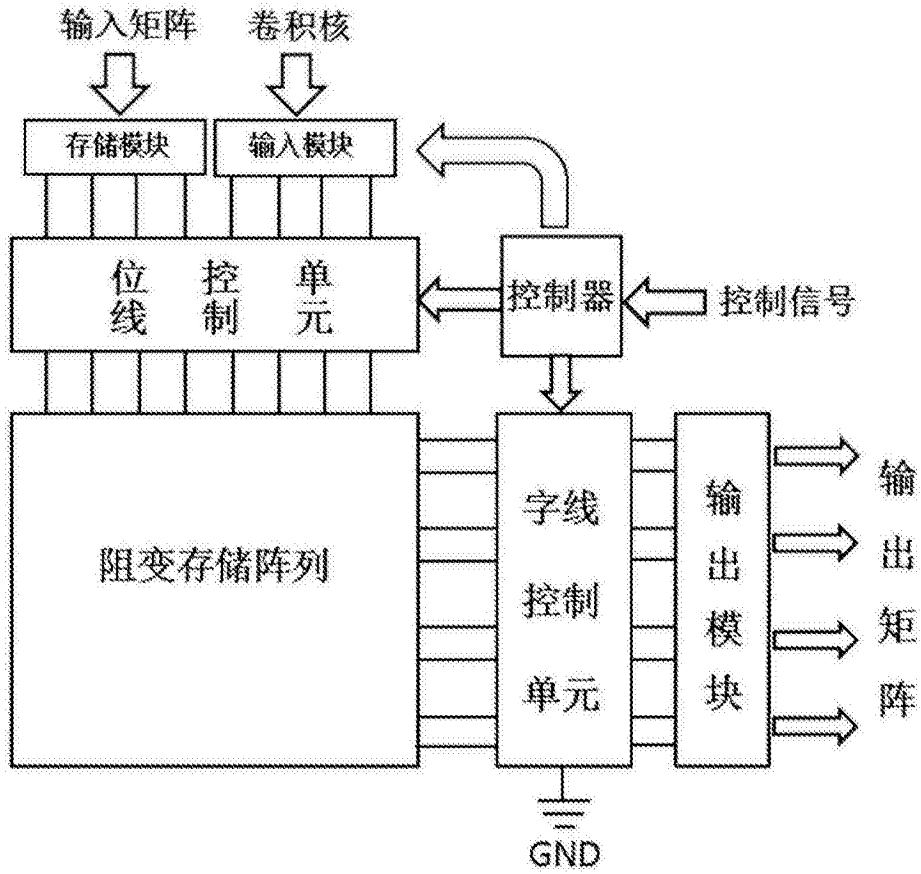


图1

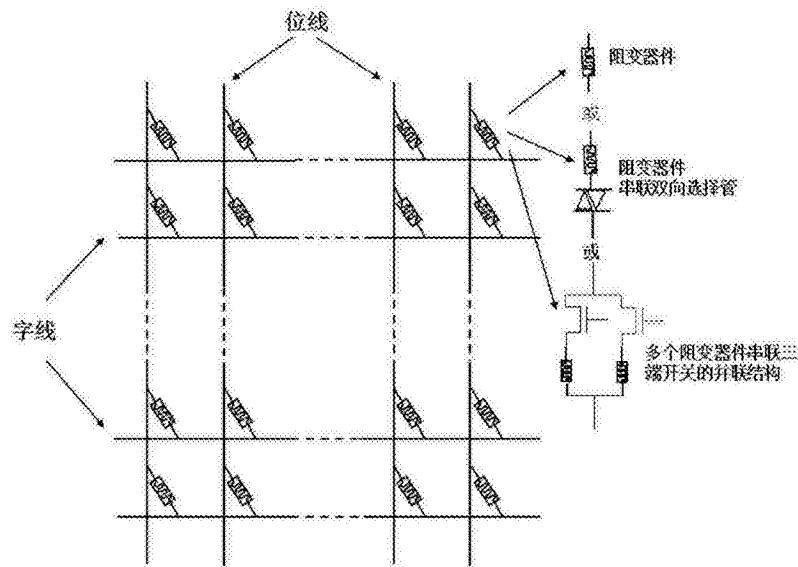


图2

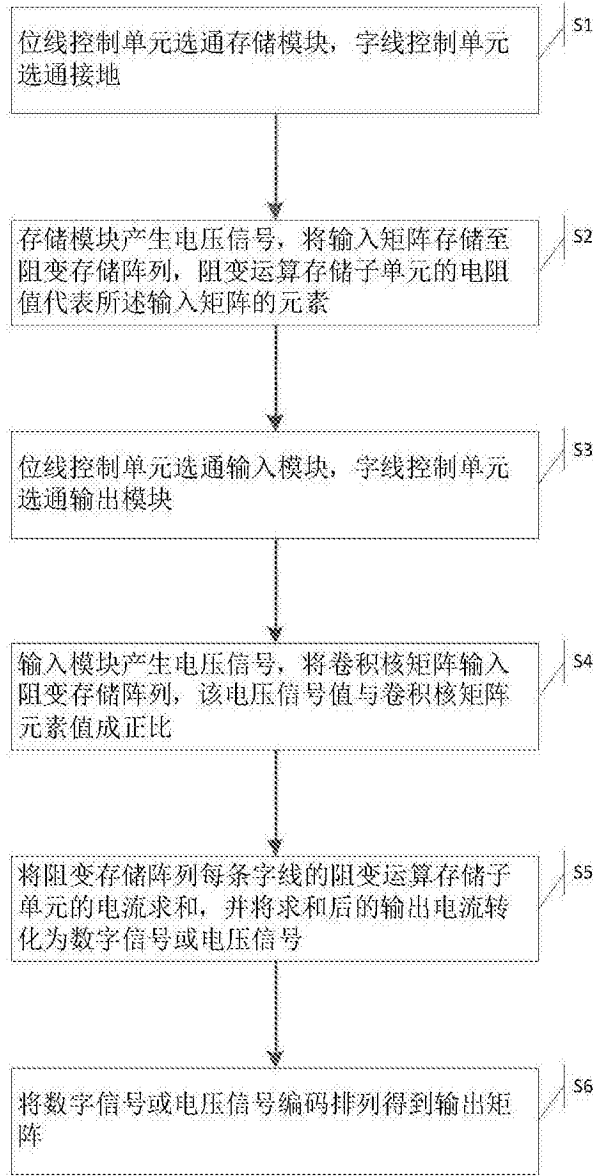


图3

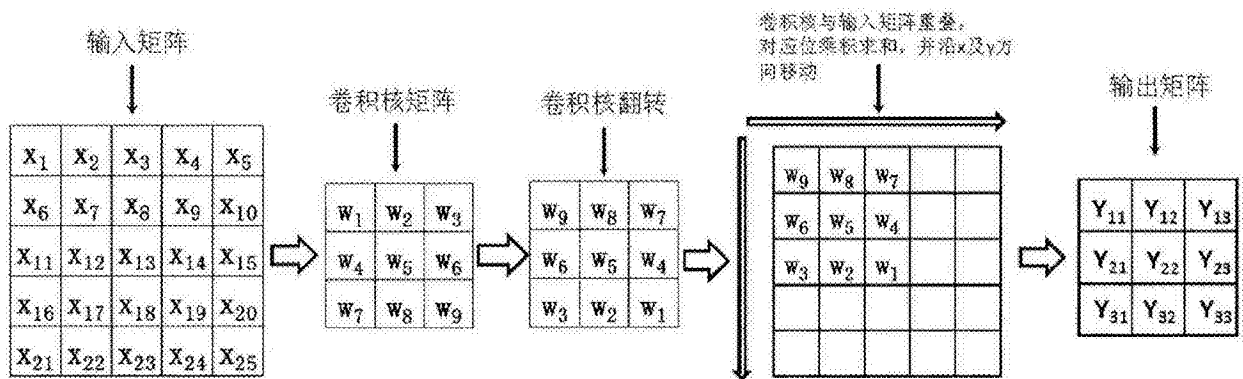


图4

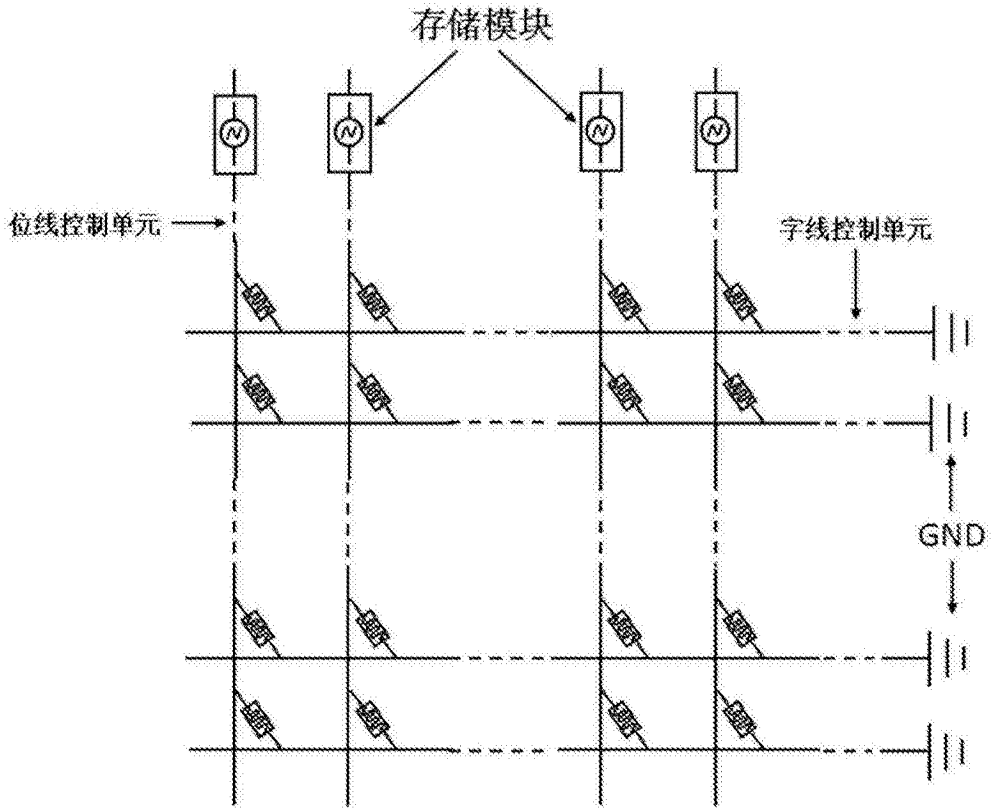


图5

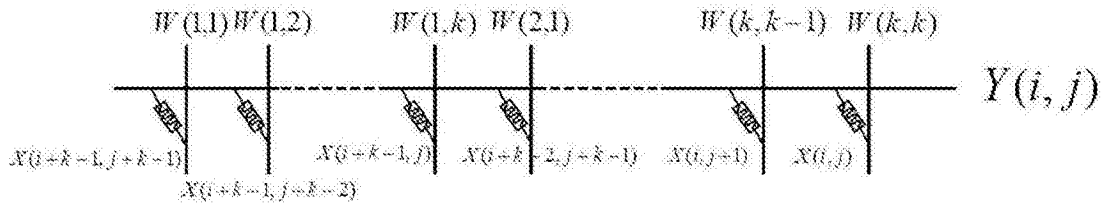


图6

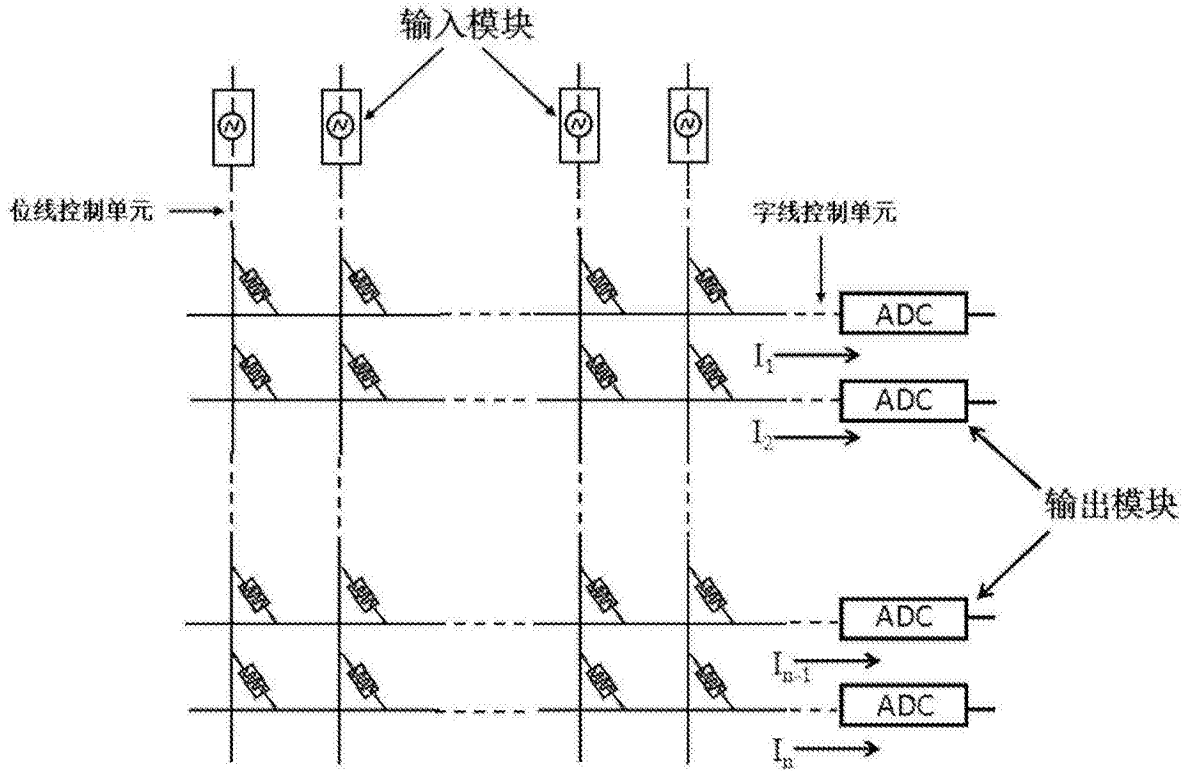


图7

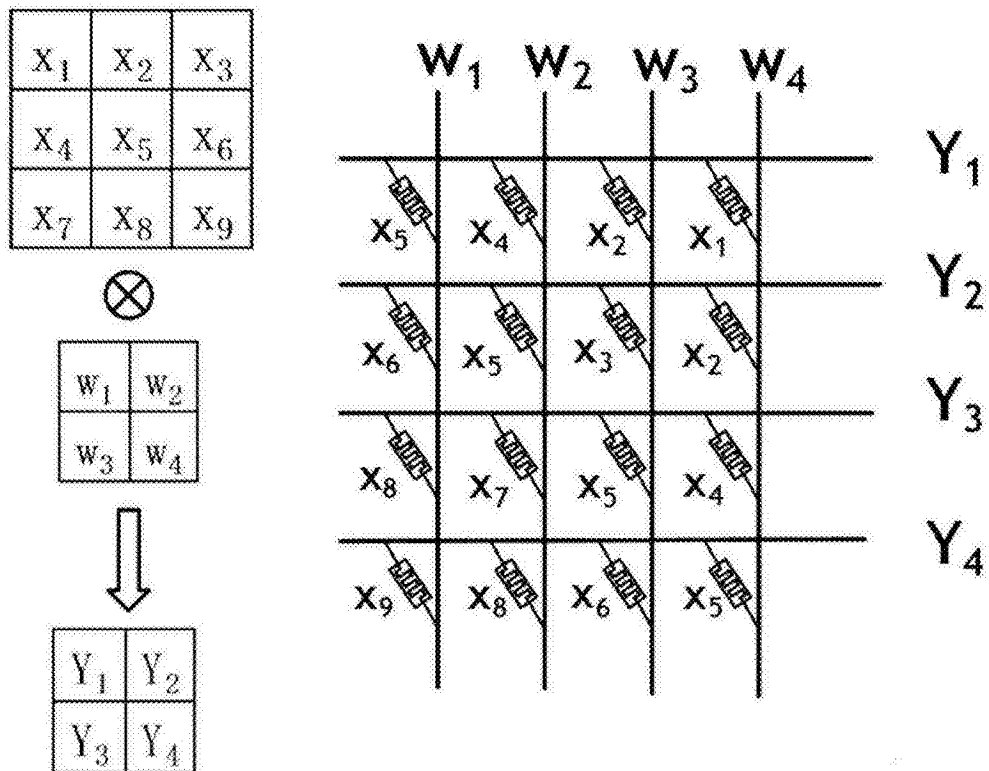


图8