



(12) 发明专利申请

(10) 申请公布号 CN 117178275 A

(43) 申请公布日 2023. 12. 05

(21) 申请号 202280028714.6

(22) 申请日 2022.04.16

(30) 优先权数据

63/175,570 2021.04.16 US

(85) PCT国际申请进入国家阶段日

2023.10.16

(86) PCT国际申请的申请数据

PCT/EP2022/060197 2022.04.16

(87) PCT国际申请的公布数据

W02022/219195 EN 2022.10.20

(71) 申请人 因纳特拉纳米系统有限公司

地址 荷兰, 赖斯韦克

(72) 发明人 A·齐亚约 S·S·库马尔

(74) 专利代理机构 北京尚诚知识产权代理有限公司 11322

专利代理师 顾小曼 熊剑

(51) Int.Cl.

G06N 3/04 (2023.01)

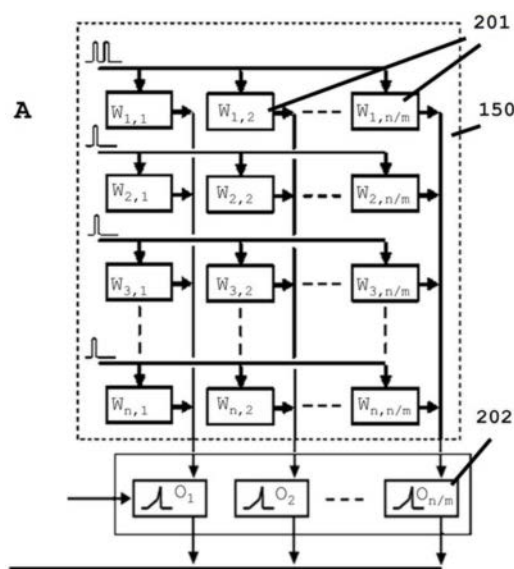
权利要求书2页 说明书10页 附图8页

(54) 发明名称

分级可重配置的多区段尖峰神经网络

(57) 摘要

本发明公开一种神经突触阵列,其包括多个尖峰神经元及多个突触元件,该多个突触元件将所述尖峰神经元互连以形成至少部分地以硬件实施的神经网络。该神经突触阵列包括权重块及输出块,每一权重块包括所述突触元件中的一个或多个,且每一输出块包括所述神经元中的一个或多个及神经元切换电路。每一输出块可经由该神经元切换电路电连接至权重块子集;该神经元切换电路配置为将包括于该权重块子集内的至少一个突触元件选择性地电连接至包括于相应输出块内的至少一个神经元,以实现将该神经突触阵列分割成电连接至所选突触元件的神经元集合。



1. 一种神经突触阵列,包括多个尖峰神经元及多个突触元件,所述多个突触元件将所述尖峰神经元互连以形成至少部分地以硬件实施的尖峰神经网络;

其中,每一突触元件配置为自多个输入中的至少一个接收突触输入信号,且适用于将权重施加至所述突触输入信号以产生突触输出信号,所述突触元件能够配置为调整由每一突触元件施加的权重;

其中,所述尖峰神经元中的每一个配置为自所述突触元件中的一个或多个接收所述突触输出信号中的一个或多个,且适用于回应于所接收到的一个或多个突触输出信号产生时空尖峰序列输出信号;

其中,所述神经突触阵列包括权重块及输出块,每一权重块包括所述突触元件中的一个或多个,且每一输出块包括所述神经元中的一个或多个以及神经元切换电路;

其中,每一输出块能够经由所述神经元切换电路电连接至权重块子集,并且其中,所述神经元切换电路配置为将包括于所述权重块子集内的至少一个突触元件选择性地电连接至包括于相应输出块内的至少一个神经元,以实现将所述神经突触阵列分割成电连接至所选突触元件的神经元集合。

2. 如权利要求1所述的神经突触阵列,其中与输出块电连接的权重块子集形成突触元件阵列内的一列或多列,和/或其中包括于特定权重块中的突触元件设置于所述神经突触阵列的相同行内,和/或其中每一输出块连接至一列权重块。

3. 如权利要求1或2所述的神经突触阵列,其中所述神经元切换电路中的每一个包括切换信号路径,所述切换信号路径包括在所述神经元切换电路的逻辑电路中实施的导线,其中所述切换信号路径配置为优选通过使用晶体管闸极能够在不同配置之间切换,其中每一配置确定包括于所述权重块子集内的哪个至少一个突触元件电连接至包括于所述输出块内的哪个至少一个神经元。

4. 如权利要求3所述的神经突触阵列,其中所述神经元切换电路配置为动态地重配置所述切换信号路径,优选地,其中动态重配置基于结合所述神经突触阵列的约束驱动分割及分段的映射方法,优选地,其中所述分段基于所述权重块和输出块大小与输入信噪比的匹配。

5. 如前述权利要求中任一项所述的神经突触阵列,其中基于一个或多个学习规则、学习速率和/或(后)可塑性机制来执行所述神经突触阵列的分段,使得神经突触子阵列中的至少两个在所述一个或多个学习规则、学习速率和/或(后)可塑性机制方面不同。

6. 如前述权利要求中任一项所述的神经突触阵列,其中所述权重块中的至少一个组织为交错结构,以便促进所述神经突触阵列内的突触元件的切换和/或组合,和/或其中对于每一输出块,所述切换为独立可控的,以便获得较高映射灵活性,和/或其中所述输出块中的至少一个组织为交错结构,和/或其中将所述输出块中的任一个的神经元输出广播至一个或多个神经突触子阵列。

7. 如前述权利要求中任一项所述的神经突触阵列,其中所述输出块中的一个内的神经元中的每一个随着电位的分级模拟变化而被动地传导。

8. 如前述权利要求中任一项所述的神经突触阵列,其中所述神经突触阵列分段成神经突触子阵列,所述神经突触子阵列能够分离并且具有其自身的必需电路,和/或所述神经突触阵列配置为处理单一模态且针对所述单一模态优化。

9. 如权利要求8所述的神经突触阵列,其中不同神经突触子阵列的神经元之间存在长程及短程互连,并且其中近端神经突触子阵列之间存在比远端神经突触子阵列之间更密集的连接性。

10. 如前述权利要求中任一项所述的神经突触阵列,其中所述输出块中的至少一个为可受控的,因为所述输出块中的所述至少一个内的特定神经元的累积周期和/或积分常数能够经由控制信号来控制。

11. 如前述权利要求中任一项所述的神经突触阵列,其中所述神经元中的一个的神经元薄膜电位在模拟域中实施为电容器两端的电压,或实施为储存于数字锁存器中的多位变量;或在数字域中使用CMOS逻辑电路实施。

12. 一种尖峰神经处理器,包括:

数据至尖峰编码器,其将数字或模拟数据编码为尖峰;

如前述权利要求中任一项所述的神经突触阵列,其配置为将由所述数据至尖峰编码器输出的尖峰视为输入,且配置为输出时空尖峰序列作为所述输入的结果;以及

尖峰译码器,其配置为对源自所述神经突触阵列的时空尖峰序列进行译码。

13. 一种用于对神经突触阵列进行配置的方法,其中:

所述神经突触阵列包括多个尖峰神经元及多个突触元件,所述多个突触元件将所述尖峰神经元互连以形成至少部分地以硬件实施的网络;

其中,每一突触元件适用于自多个输入中的至少一个接收突触输入信号,且将权重施加至所述突触输入信号以产生突触输出信号,所述突触元件能够配置为调整由每一突触元件施加的权重;

其中,所述尖峰神经元中的每一个适用于自所述突触元件中的一个或多个接收所述突触输出信号中的一个或多个,且回应于所接收到的一个或多个突触输出信号产生时空尖峰序列输出信号;

其中,所述方法包括:

将所述神经突触阵列划分为权重块及输出块,每一权重块包括所述突触元件中的一个或多个,且每一输出块包括所述神经元中的一个或多个以及神经元切换电路;

使每一输出块能够经由所述神经元切换电路电连接至权重块子集;

对所述神经元切换电路进行配置以将包括于所述权重块子集内的至少一个突触元件选择性地电连接至包括于相应输出块内的至少一个神经元,以实现将所述神经突触阵列分割成电连接至所选突触元件的神经元集合。

14. 如权利要求14所述的方法,其中与输出块电连接的权重块子集形成所述突触元件阵列内的一列或多列,和/或其中包括于特定权重块中的突触元件设置于所述神经突触阵列的相同行内。

15. 如权利要求13或14所述的方法,其中所述神经元切换电路中的每一个包括切换信号路径,所述切换信号路径包括在所述神经元切换电路的逻辑电路中实施的导线,其中所述切换信号路径配置为优选通过使用晶体管闸极能够在不同配置之间切换,其中每一配置确定包括于所述权重块子集内的哪个至少一个突触元件电连接至包括于所述输出块内的哪个至少一个神经元。

分级可重配置的多区段尖峰神经网络

技术领域

[0001] 本发明涉及自动信号识别技术,且更具体而言,涉及一种用于尖峰神经元的分级可重配置的多区段网络的系统及方法。

背景技术

[0002] 深度神经网络(DNN)构成大量机器学习应用的基础;自语音及图像识别开始,使用DNN的应用的数目呈指数增长。最初,硬件(深度网络)加速器已在标准同步数字逻辑上实施。神经网络的高度并行性在数字系统的(通常)串行及时间多任务处理中无法复制;相反,硬件DNN仿真器的计算基元作为模拟计算节点实现,其中内存及处理元件经共定位,在速度、大小及功率消耗方面提供显著改良。

[0003] 在生物神经网络模型中,每一个别神经元异步地且经由稀疏事件或尖峰通信。在此基于事件的尖峰神经网络(SNN)中,仅改变状态的神经元产生尖峰且可触发后续层中的信号处理,因此节省计算资源。具体而言,SNN结合异步分布式架构,从而借助于局部尖峰驱动计算、局部或整体反馈及在线学习来处理稀疏二进制时间序列。类似于生物神经处理系统,基于混合信号的SNN处理器保存了两个其基本特性:时间的外显表示及空间的外显使用,从而将每一神经元/突触元件的专用物理电路实例化。

[0004] 本质上,此SNN实施使用混合式模拟-数字信号表示,即脉冲/尖峰的序列在事件的时间序中传输模拟信息,所述模拟信息在神经元的树突(输入)中转换回模拟信号。信息由神经元的群体中发生的活动的模式编码,且突触(至后续神经元的连接)可根据其接收的脉冲来调适其功能,从而提供信号传输能效以及储存及再调用信息的灵活性。SNN可直接应用于模式识别及传感器数据融合,此依赖于可将振幅域、时域及频域特征编码成独特空间及时间编码的尖峰序列的原理。这些序列的产生依赖于尖峰神经网络的一个或多个区段的使用。

[0005] 目前已知的用于将SNN映像至硬件基板上的方法在例如用于SNN内的特定功能的神经元的数目、经由突触的连接性及这些神经元及突触的配置方面不允许有较大灵活性。

[0006] 此外,尽管通过模拟电子电路系统来极高效地实现SNN所需的基本操作,但制造不匹配会诱导其功能性质的失真。参见M.J.M.Pelgrom、A.C.J.Duinmaijer、A.P.G.Welbers的“Matching properties of MOS transistors(MOS晶体管的匹配性质)”,IEEE Journal of Solid-State Circuits,vol.24,no.5,pp.1433-1439,1989(IEEE固态电路杂志,第24卷,第5期,第1433-1439页,1989年)。尤其在较小工艺几何形状及较低工作电流下,这些电路越来越容易受到量子效应及外来噪声的影响,从而有效地降低信噪比且限制处理性能。参见K.L.Shepard、V.Narayanan的“Noise in deep submicron digital design(深次微米级数字设计中的噪声)”,IEEE International Conference on Computer-Aided Design,pp.524-531,1997(IEEE计算机辅助设计国际会议,第524-531页,1997年)。

[0007] 在其中驱动器、偏置以及神经元及突触由更多数目的装置经由长时间互连而共享的大型阵列的情况下,这些非理想性的影响增加。

发明内容

[0008] 本发明中所公开的神突触区段组织为突触电路及神经元单元的重复阵列。本发明涵盖这些阵列如何对区段进行分割及映像的方式,且我们描述神经突触阵列及随附连接可重配置及分段的方法及机制,随后在将SNN映像至硬件基板上提供较高灵活性。映像方法结合区段的约束驱动分割及映像(根据各区段的大小(神经元的数目)、其连接性(拓扑及突触数)及其配置(权重及层数)),其中所选定义及限定约束为与区段的网络的指定功能相关的性能度量。在本公开的其余部分中详述的此方法产生一种用于实现能够对复杂感测信号进行模式识别的尖峰神经元的大型网络的有效方式。

[0009] 根据本公开的第一方面,公开一种神经突触阵列,其包括多个尖峰神经元及多个突触元件,该多个突触元件将尖峰神经元互连以形成至少部分地以硬件实施的尖峰神经网络。每一突触元件配置为自多个输入中的至少一个接收突触输入信号,且适用于将权重施加至突触输入信号以产生突触输出信号,所述突触元件配置为调整由每一突触元件施加的权重。此外,尖峰神经元中的每一个适用于自突触元件中的一个或多个接收突触输出信号中的一个或多个,且适用于回应于所接收到的一个或多个突触输出信号产生时空尖峰序列输出信号。

[0010] 神经突触阵列包括权重块及输出块,每一权重块包括突触元件中的一个或多个,且每一输出块包括神经元中的一个或多个及神经元切换电路。每一输出块可经由神经元切换电路电连接至权重块子集,且其中神经元切换电路配置为将包括于权重块子集内的至少一个突触元件选择性地电连接至包括于相应输出块内的至少一个神经元,以实现将神经突触阵列分割成电连接至所选突触元件的神经元集合。

[0011] 根据本公开的第一方面的实施例,与输出块电连接的权重块子集形成突触元件阵列内的一列或多列,和/或其中将包括于特定权重块中的突触元件设置于神经突触阵列的相同行内和/或其中每一输出块连接至一列权重块。

[0012] 根据本公开的第一方面的实施例,神经元切换电路中的每一个包括切换信号路径,所述切换信号路径包括在神经元切换电路的逻辑电路中实施的导线,其中切换信号路径配置为可优选通过使用晶体管闸极在不同配置之间切换,其中每一配置确定包括于权重块子集内的哪一至少一个突触元件电连接至包括于输出块内的哪一至少一个神经元。

[0013] 根据本公开的第一方面的实施例,神经元切换电路配置为动态地重配置切换信号路径,优选地,其中动态重配置基于结合神经突触阵列的约束驱动分割及分段的映像方法,优选地,其中分段基于权重块及输出块大小与输入噪声比的匹配。

[0014] 根据本公开的第一方面的实施例,基于一个或多个学习规则、学习速率和/或(后)可塑性机制来执行神经突触阵列的分段,使得神经突触子阵列中的至少两个在一个或多个学习规则、学习速率和/或(后)可塑性机制方面不同。

[0015] 根据本公开的第一方面的实施例,权重块中的至少一个组织为交错结构,以便促进神经突触阵列内的突触元件的切换和/或组合,和/或其中对于每一输出块,该切换为独立可受控的,以便获得较高映射灵活性,和/或其中输出块中的至少一个组织为交错结构,和/或其中将输出块中的任一个的神经元输出广播至一个或多个神经突触子阵列。

[0016] 根据本公开的第一方面的实施例,输出块中的一个内的神经元中的每一个随着电位的分级模拟变化而被动地传导。

[0017] 根据本公开的第一方面的实施例,将神经突触阵列分段成神经突触子阵列,所述神经突触子阵列为可分离的且具有其自身的必需电路,和/或配置为处理单一模态且针对单一模态优化。

[0018] 根据本公开的第一方面的实施例,不同神经突触子阵列的神经元之间存在长程及短程互连,且其中近端神经突触子阵列之间存在比远端神经突触子阵列之间更密集的连接性。

[0019] 根据本公开的第一方面的实施例,输出块中的至少一个为可受控的,因为输出块中的至少一个内的特定神经元的累积周期和/或积分常数可经由控制信号来控制。

[0020] 根据本公开的第一方面的实施例,神经元中的一个的神经元薄膜电位在模拟域中实施为电容器两端的电压,或实施为储存于数字锁存器中的多位变量;或在数字域中使用CMOS逻辑电路实施。

[0021] 根据本公开的第二方面,公开一种尖峰神经处理器。尖峰神经处理器包括:数据至尖峰编码器,其将数字或模拟数据编码为尖峰;如根据本公开的第一方面的神经突触阵列,其配置为将由数据至尖峰编码器输出的尖峰视为输入,且配置为输出一时空尖峰序列作为该输入的结果;以及尖峰译码器,其配置为译码源自神经突触阵列的时空尖峰序列。

[0022] 根据本公开的第三方面,公开一种用于对神经突触阵列进行配置的方法。神经突触阵列包括多个尖峰神经元及多个突触元件,该多个突触元件将尖峰神经元互连以形成至少部分地以硬件实施的神经网络。每一突触元件适用于自多个输入中的至少一个接收突触输入信号,且将权重施加至突触输入信号以产生突触输出信号,突触元件配置为调整由每一突触元件施加的权重。此外,尖峰神经元中的每一个适用于自突触元件中的一个或多个接收突触输出信号中的一个或多个,且回应于所接收到的一个或多个突触输出信号产生时空尖峰序列输出信号。

[0023] 该方法包括:将神经突触阵列划分为权重块及输出块,每一权重块包括突触元件中的一个或多个,且每一输出块包括神经元中的一个或多个及神经元切换电路;使每一输出块可经由神经元切换电路电连接至权重块子集;对神经元切换电路进行配置以将包括于权重块子集内的至少一个突触元件选择性地电连接至包括于相应输出块内的至少一个神经元,以实现将神经突触阵列分割成电连接至所选突触元件的神经元集合。

[0024] 根据本公开的第三方面的实施例,与输出块电连接的权重块子集形成突触元件阵列内的一列或多列,和/或其中将包括于特定权重块中的突触元件设置于神经突触阵列的相同行内。

[0025] 根据本公开的第三方面的实施例,神经元切换电路中的每一个包括切换信号路径,所述切换信号路径包括在神经元切换电路的逻辑电路中实施的导线,其中切换信号路径配置为可优选通过使用晶体管闸极在不同配置之间切换,其中每一配置确定包括于权重块子集内的哪一至少一个突触元件电连接至包括于输出块内的哪一至少一个神经元。

附图说明

[0026] 现在将参考随附示意图而仅借助于实例来描述实施例,在附图中,对应附图标记指示对应部分,且在所述附图中:

[0027] 图1展示具有 n^2 个突触的正方形神经突触阵列的示意图;

[0028] 图2A到图2C展示可重配置神经突触阵列的组织示意图:在图2A中,主要组织单元为分别收集 m 个权重及 m 个神经元的权重块 $w_{i,j}$ 及输出块 O_j ;在图2B中,展示权重块 $w_{i,j}$ 的结构;在图2C中,展示输出块 O_j 的结构;

[0029] 图2D展示列/神经元译码器中的神经元切换电路的可能配置的例示性实施例;

[0030] 图3示意性地展示如何将例示性可重配置阵列(如例如图2中所展示)划分为多个区段/子阵列;

[0031] 图4A到图4D示意性地展示可重配置神经突触阵列的例示性组织:在图4A中,为 n^2 个突触及 n 个神经元;在图4B中,将图4A中的阵列划分为两个区段,每一区段具有 $n^2/2$ 个突触及 $2 \times n$ 个神经元;在图4C中,将阵列划分为四个区段,每一区段具有 $n^2/4$ 个突触及 $4 \times n$ 个神经元;以及在图4D中,将阵列划分为八个区段,每一区段具有 $n^2/8$ 个突触及 $8 \times n$ 个神经元;

[0032] 图5展示可重配置神经突触阵列的组织示意图:区段(子阵列)中的每一个可具有不同的每神经元突触数;

[0033] 图6A展示结合数据至尖峰编码器、多区段模拟/混合信号神经突触阵列及尖峰译码器的尖峰神经处理器的例示性图示;

[0034] 图6B展示互连结构的示意图。

[0035] 附图仅意谓出于说明的目的,且不充当对如由权利要求列出的范围或保护的限制。

具体实施方式

[0036] 在下文中,将进一步详细地描述某些实施例。然而,应了解,这些实施例不可被视为限制本公开的保护范围。

[0037] 图1展示具有 n^2 个突触及 n 个神经元的正方形神经突触阵列100的示意图。神经突触阵列100包括具有突触101的突触矩阵150及具有神经元103的神经元行160。突触及神经元可部分地或完全地以硬件实施。

[0038] 突触矩阵150中的突触101中的每一个具有归因于其的某一权重 $w_{i,j}$,其中 i 表示突触矩阵150的第 i 行,且 j 表示突触矩阵150的第 j 列。突触矩阵150的每一行由相应输入102驱动。输入102为电信号且可为例如电压或电流尖峰(序列)。突触矩阵150的每一列 j 驱动表示为 N_j 的相应输出神经元103。

[0039] 权重 $w_{i,j}$ 中的每一个表示至突触矩阵150的第 i 行的相应输入102与连接至突触矩阵150的第 j 列的神经元 N_j 之间的连接的强度或幅值。换言之,若权重 $w_{i,j}$ 增加,则至第 i 行的相应输入102与连接至突触矩阵150的第 j 列的神经元 N_j 之间的连接加强。

[0040] 当特定神经元 N_j 的薄膜膜电位(其与神经元的薄膜电荷相关)达到比值(称作“阈值”)时,神经元激发。此电信号表示为电压或电流尖峰,且经由尖峰输出线180发送至SNN的另一部分。

[0041] 可控制神经元行160,因为例如可经由控制信号170来控制特定神经元103的累积周期和/或积分常数(尽管仅展示一个控制信号170,但每一神经元可存在一个或多个单独控制信号)。以此方式,可控制神经元103对来自突触矩阵150的信号的反应。

[0042] 此正方形神经突触阵列为尖峰神经网络硬件的实例,诸如在US 63/107,498中描

述的尖峰神经网络硬件,其利用尖峰神经元、突触的可配置阵列,所述阵列使用有助于实现任何任意连接拓朴的可程序化互连结构进行连接。此处,使用分布式组件来描述有效神经突触阵列,该分布式组件由驱动 n^2 个突触及 n 个神经元的 n 个输入构成,即该系统在阵列的接合点处结合电突触的存在,同时阵列的周边包含多行神经元电路,其仿真生物神经元的胞体及轴突丘的作用。

[0043] 此阵列可对区段的阵列进行分割及映像,所述区段中的每一个含有尖峰神经元的可程序化网络。

[0044] 图2A到图2C展示包括 n^2 个突触元件及 n 个神经元的可重配置神经突触阵列的组织示意图。在图2A中,主要组织单元为分别收集 $1 \times m$ 个突触元件及 m 个神经元的权重块201及输出块202。

[0045] 图2B展示具有输入线的单一权重块201。为了支持更大扇入,可将在多个突触列上散布的多个突触元件的突触输出引导至单一神经元中,从而有效地增加神经元对 $n \times m$ 个独特输入的潜在扇入。此通过将突触收集至权重块中来达成。可重配置神经突触阵列的组织描述由权重块内的 m 数目个突触元件控制的可能架构的空间,其中 $m \in [1, n]$,其中神经元的最大扇入为 $n \times m$,输入端口的数目为 $n \times m$,且经由单一输入尖峰所广播的神经元的数目为 n/m 。当所有映像神经元之间未共享输入时,增加 $W_{n, n/m}$ 数目个权重块使突触利用率更高。

[0046] 因此,权重块201中的突触列(多列突触元件)的数目等于 m ,如图2B中所展示,且 n/m 为神经突触阵列150的每一行中的权重块201的总数目,其中 n 为突触矩阵150中的总列数(假定为正方形阵列)。权重块可包括多行突触元件以及多列突触元件,然而在这一实例中,每一权重块201中的行数为1。每一权重块表示为 $W_{i,j}$,其中 $i \in [1, n]$ 且 $j \in [1, n/m]$ 。

[0047] 应注意,尽管为了简单起见,图2A为权重块201的每一行描绘单一输入线,但权重块201的每一行实际上接收多个输入,例如,包括 m 个突触元件的权重块201可具有 m 个输入,如图2B中所展示,使得每一行中的权重块201连接到 m 条输入线。类似地,为了简单起见,图2A描绘用于传输权重块201的每一列的输出的单一线。然而,权重块201的每一列实际上产生多个输出,例如包括 m 个突触元件的权重块201可产生 m 个输出,如图2B中所展示,使得每一列中的权重块201连接到 m 条输出线。

[0048] 图2C展示单一输出块202。由一列权重块201驱动的每一输出块202中的神经元数目为 m 。每一输出块表示为 O_i ,其中 $i \in [1, n/m]$ 。

[0049] 可因此将正方形可重配置神经突触阵列划分为大小为 m 的 n/m 个输出块,其中 $m \leq n$ 。因此形成可分离神经突触阵列,其具有 n 个独特输入端口及包括 m 个神经元的 n/m 个输出块。

[0050] 独特输入的总数目由此增加,且可用于增加个别神经元的扇入和/或以便为尖峰分布提供更多粒度。反馈连接与阵列一起缓慢地缩放以用于有效映像循环及多层网络。每行的权重块的数目 n/m 因此表明 n^2 个神经突触阵列的内部分割;总体而言,独特输入的数目增加至 $n \times m$ 。若 $m = n$,则每一突触均具有其自身专用的输入端口,而对于 $m = 1$,获得原始的 n^2 个阵列,因此在所有神经元之间共享输入。

[0051] 在图2B中,针对可重配置神经突触阵列的组织表示单一权重块 $W_{i,j}$ 的结构。

[0052] $W_{n, n/m}$ 权重块中的每一个遵循阵列组织,即沿着一行权重块广播 m 个输入的集合(表示为 in^i ,其中 $i \in [1, n]$),亦即将单一输入 in_x^i (其中 $x \in [1, m]$)直接分布至 n/m 个神经元。

[0053] 权重块 $W_{i,j}$ (其中 $i \in [1, n]$ 且 $j \in [1, n/m]$)的第 x 个突触元件 (其中 $x \in [1, m]$) 表示为 $w_{i,j}^x$ 。沿着突触元件的第 i 行广播 m 个输入的集合 (表示为 in^i , 其中 $i \in [1, n]$) ; 亦即, 将 m 个输入的集合中的单一输入 in_x^i (其中 $x \in [1, m]$) 分布至 n/m 个神经元。

[0054] 在图2C中, 针对可重配置神经突触阵列的组织展示输出块 O_j 的结构。

[0055] 一系列权重块201包括可连接至输出块202的多个突触列, 且将来自突触列的输出252传递至列/神经元译码器250中。列/神经元译码器250包括用于将突触列输出252连接至输出块202的一组或多组神经元254的切换电路260, 切换电路260由一个或多个选择信号251控制。已将离开列/神经元译码器250的切换信号切换至输出块 O_j (其中 $j \in [1, n/m]$) 的一个或多个神经元254 (其由 N_x^j 表示, 其中 $x \in [1, m]$) , 从而使得能够将突触列输出252连接至神经元或神经元集合的不同组合。

[0056] 选择信号251可用于选择列/神经元译码器250如何将突触列输出252自输出块202的神经元254的集合切换至神经元254的另一集合。一个或多个神经元控制信号253可用于确定神经元集合254或个别神经元的操作性质。

[0057] 在图2D中, 展示当四个突触列270充当至列/神经元译码器250的输入时, 列/神经元译码器250中的神经元切换电路260的可能配置的例示性实施例。在此实例中, 列/神经元译码器250经布置以向四个相邻的突触列270的每一群组提供如所展示的四种可能的输出神经元配置280, 即四个突触列可使其输出选择性地针对输出块202的神经元254的四个不同集合。列/神经元译码器250可与不同数目个突触列270及输出神经元254连接, 且突触列及输出神经元的数目可不同, 且输出神经元配置280的数目及布置可不同。

[0058] 接下来论述神经元切换电路260的四种配置280。在第一配置中, 四个突触列270经由切换信号路径261充当至四个输出神经元的相应输入。在第二配置中, 四个突触列270中的两对经由切换信号路径261充当至两组输出神经元的输入, 即突触列 i_0 及 i_1 伺服输出神经元 o_0 且突触列 i_2 及 i_3 伺服输出神经元 o_2 。在第三配置中, 三个突触列 i_0 、 i_1 及 i_2 伺服一个输出神经元集合 o_0 , 且其余突触列 i_3 经由切换信号路径261伺服输出神经元 o_3 。在第四配置中, 所有四个突触列均经由切换信号路径261伺服单一输出神经元集合 o_0 。

[0059] 切换信号路径261可包括实施于切换电路260的逻辑电路中的导电路径, 且其在特定实施中可通过例如可配置或固定晶体管闸极在不同配置之间切换。可在神经突触阵列的初始制造时 (例如在出厂时) 设置切换信号路径261的特定连接, 和/或可在运行时经由列/神经元译码器250和/或神经元切换电路260的程序化或其他配置来设置。神经元切换电路260可例如在神经突触阵列的操作期间动态地对切换信号路径261进行配置。如所提及, 一个或多个选择信号251可用于选择列/神经元译码器250如何自神经元254的集合切换至神经元254的另一集合。动态重配置可基于结合区段的约束驱动分割及映像的特定映像方法, 下文将进一步描述其例示性实施例。举例而言, 动态重配置可基于神经突触阵列中的特定区段的信噪比S/N来进行。此将在下文进一步解释。

[0060] 神经元切换电路260因此确定哪些突触列270向哪些输出神经元集合280发送输出。换言之, 神经元切换电路260因此实施——作为神经元切换电路260的当前作用中配置的结果——神经突触阵列的特定分段。

[0061] 为了支持较大扇入, 权重块 $W_{n,n/m}$ 可组织为交错结构以促进 m 个相邻突触列的切换/

组合。可独立地控制用于 n/m 次神经元切换中的每一集合的配置以用于更多映射灵活性。对于循环尖峰或层间连接的潜时敏感布线 (latency-sensitive routing), 可考虑若干选项, 例如利用神经元群组或集合的交错组织在输出神经元 i 与所选输入之间提供直接路径, 使得输入可快速布线至任何群组, 从而使尖峰可挠地注射至阵列中, 其中神经元输出可自动地广播至个别或多个群组。

[0062] 对阵列进行分段能够实现性能的鲁棒性及可靠度方面的均一性, 即分段允许: (i) 鲁棒的功率分布 (竖直的宽金属总线 (用于降低电阻) 可置放于突触行边界处以缓解IR压降问题); (ii) 若对权重保持电路进行存取的接口需要时钟, 则分段创建时钟回退策略所需的层次结构; (iii) 分段提高信号完整性——而非使单一预突触具有较大驱动强度, 分段允许分布式方法, 因为信号可在每次分段后经回退/放大; (iv) 分段提高阵列内的可观察性且简化验证过程。

[0063] 图3示意性地展示如何将例示性可重配置阵列 (如例如图2中所展示) 划分为多个区段/子阵列。

[0064] 第一子阵列301包括 $i \in [1, n]$ 且 $j \in [1, x]$ 的权重块 W_{ij} 。第二子阵列302包括 $i \in [1, n]$ 且 $j \in [x+1, y]$ 的权重块 W_{ij} 。最终子阵列303包括 $i \in [1, n]$ 且 $j \in [z+1, n/m]$ 的权重块 W_{ij} 。此处, x, y, z 为自然数且 $x \leq y \leq z \leq n/m$ 。可存在更多子阵列。若干 n 条输入线304馈送不同的权重块。权重块 W_{ij} 中的每一个连接到输出块 O_j 。每一子阵列可包括分别分段为子阵列301、302、303的输出块305、306、307的若干输出块。当特定神经元 N_j 的薄膜膜电位 (其与神经元的薄膜电荷相关) 达到比值/具体值时, 神经元激发。此电信号表示为电压或电流尖峰, 且经由尖峰输出线309发送至SNN的另一部分。

[0065] 可控制输出块中的每一个, 因为例如特定神经元的累积周期和/或积分常数可经由控制信号308控制。以此方式, 可控制神经元对来自子阵列301、302、303中的权重块的信号的反应。

[0066] 如先前所提及, 通过模拟电子电路系统来极高效地实现SNN所需的基本操作, 然而, 制造不匹配会诱导其功能性质的失真。尤其在较小工艺几何形状及较低工作电流下, 这些电路越来越容易受到量子效应及外来噪声的影响, 此有效地降低信噪比且限制处理性能。在其中驱动器、偏置以及神经元及突触由更多数目的装置经由长时间互连而共享的大型阵列的情况下, 这些非理想性的影响增加。若映像方法结合区段的约束驱动分段及映像, 则将大型阵列分割成较小区段 (子阵列) (每一区段均具有自身的必需电路) 使得实现性能均一性且减少这些非理想性。

[0067] 在一实施例中, 可基于目标信号处理功能的特性 (例如增益、滤波、相乘、相加) 或操作的频率/时间常数来组织区段。

[0068] 在另一实施例中, 可基于系统的目标应用案例、预处理信号序列/路径中的信道数目、所关注的频带、特征集合的复杂度及特性, 或目标信噪比对神经突触阵列的不同区域/区段进行配置。

[0069] 在另一实施例中, 可基于受生物物理/化学启发的定义 (例如时空 (长期及短期) 评估、老化假影、大脑通信成本及区域/层) 来组织区段。

[0070] 在另一实施例中, 可基于学习规则、学习速率及相关联的基于成本的机制或 (后) 可塑性机制, 或基于预定义的基于性能质量因子的定义来定义区段。

[0071] 在另一实施例中,可基于稳态调节(局部及整体)需求、或误差传播/校准能力及机制限制、或限制机制(诸如兴奋/抑制比、串音或网络饱和机制)来对神经突触阵列内的区段进行配置。

[0072] 图4A至图4D示意性地展示可重配置神经突触阵列400的例示性组织:在图4A中,为 n^2 个突触401及 n 个神经元402;在图4B中,将图4A中的阵列400划分为两个区段,每一区段具有 $n^2/2$ 个突触及 $2 \times n$ 个神经元;在图4C中,将阵列400划分为四个区段,每一区段具有 $n^2/4$ 个突触401及 $4 \times n$ 个神经元402;且在图4D中,将阵列400划分为八个区段,每一区段具有 $n^2/8$ 个突触401及 $8 \times n$ 个神经元402。

[0073] 神经突触区段可由驱动 n^2 个突触及 n 个神经元的 n 个输入构成(如在图1到图3、图4A中)。通过减小树突/阵列距离,复合物及隔室缩短延迟且较低噪声,即为了减少功率消耗且增加相同区域的每突触连接的神经元数,可将图4A中的阵列细分成两个区段(图4B),每一区段具有 $n^2/2$ 个突触及 n 个神经元;因此,两个区段当中的总突触数保持相同,而神经元的数目加倍。类似地,将图4A中的阵列划分为四个区段(图4C)(每一区段具有 $n^2/4$ 个突触)可使总神经元数增加四倍;或相应地,将该阵列划分为八个区段(图4D)(每一区段具有 $n^2/8$ 个突触)可使神经元数增加八倍。

[0074] 阵列的信息容量取决于信(S)杂(N)比 S/N 且增加为 $\log_2(1+S/N)$ 。然而,经由阵列传递信号的能量成本增加为 $\sqrt{S/N}$ 。因此,随着 $\sqrt{S/N}$ 增加,阵列的效率下降,即所有组件均试图传输相同信号,尽管每单位固定成本的信息增加。当这两种竞争趋势达到平衡时,就会出现最优状态。因此,固定成本与信令成本的比率越高,最优阵列就越大。

[0075] 考虑到在组件当中分布所分配 S/N 以最大化整个系统的性能,最优阵列大小亦取决于电路的其他部分的成本。

[0076] 图5展示可重配置神经突触阵列500的组织的示意图:区段(子阵列)501中的每一个可具有不同的每神经元突触数。在区段501层级上,为了受益于传感器输出的准确度增加和/或区分特征选择及提取能力的增加,可定义各自致力于处理单一模态的专用区段。

[0077] 区段中的每一个可具有不同的每神经元突触数,其中 $m = [1, .n]$,根据本发明的描述定义。

[0078] 输入的 S/N 极大地影响阵列的最优大小,因为输入噪声强加待由阵列的 S/N 接近的界限。此减小大型阵列在低输入 S/N 下的功效及最有效阵列的大小,即由于具有低 S/N 的输入含有较少信息,且较小阵列具有较低信息容量,因此最优阵列使其容量与其输入相匹配。阵列大小与输入 S/N 的匹配遵循对称形态构成(symmorphosis)的原理,以匹配系统内的容量,以便获得最优解决方案。参见E.R.Weibel, "Symmorphosis: On form and function in shaping life (对称形态构成:论塑造生命的形式及功能)," Cambridge, MA: Harvard University Press, 2000 (马萨诸塞州剑桥:哈佛大学出版社,2000年)。

[0079] 为了节省区段内的神经元数目(但防止一个组件以次优方式完成多项任务),应增强神经元内的电路的计算能力,即将更多神经受体(AMPA、NMDA、GABA)或基于神经元电导(Na、Ca、K)的定义用于神经元的特定模态,且因此改良神经元的灵敏度及信噪比。这些神经元可使用内部(神经受体或基于电导的)电路来执行功能,此通常需要若干神经元的电路。

[0080] 神经调节物质可广泛地广播但仍在局部起作用,且具体而言,仅影响具有适当神经受体的神经元。神经调节物质的作用范围进一步增加,因为其受体多样化成耦合至不同

区段内传信网络的多种子类型。因此,较小、有针对性的调整可再调谐及重配置整个网络,从而实现高效处理模块性且在需要其他信号处理功能性时重新排程任务。神经受体改变电流的电压依赖性,且随后影响活化及失活曲线的形状。

[0081] 每一受体代表多(跨)电导通道,其仿真相关联的非线性特性。NMDA受体可对突触权重提供活性依赖性修改,而AMPA受体可使快速突触电流驱动胞体。快速受体修改可充当突触功效中的活性依赖性变化的控制机制。若树突尖峰的群组(突发)足以超过特定神经元的阈值,则轴突将产生动作电位;将随后尖峰馈送回至树突中,且与胞体信号一起相乘且与NMDA受体信号相加,以随后产生权重控制。

[0082] 大体而言,反馈信号可为那些可用信号的子集,且可作用于系统参数的子集(即,在补偿扰动时保留特定的固有性质)。参见T.O' Leary、A.H.Williams、A.France、E.Marder, "Cell types, network homeostasis, and pathological compensation from a biologically plausible ion channel expression model (细胞类型、网络稳态及来自在生物学上似乎合理的离子信道表现模型的病理性补偿)," *Neuron*, vol. 82, pp. 809-821, 2014 (神经元, 第82卷, 第809-821页, 2014年)。

[0083] 当网络扰动发生时,神经元计算组件基于扰动的符号、强度及相位产生尖峰。若神经元丛集的共振响应扭曲或改变,则每一神经元均调整尖峰的相位。随后,神经元计算组件可执行与时间依赖计算。

[0084] 在足够高水平的异质性下(例如神经突触参数,如积分器电容、发射阈值、不应期中的异质性增加),网络表明神经活性的空间无序模式。若网络活动特征曲线较窄,则增加对由网络编码的空间刺激的分布的均质性的需要。

[0085] 图6A展示尖峰神经处理器600的例示性图示,该尖峰神经处理器600结合一个或多个数据至尖峰编码器606、多区段模拟/混合信号神经突触阵列607及尖峰译码器611。

[0086] 数据用作输入模块(诸如AER 601)、接口(诸如例如SPI 602、I²C 603和/或I/O(输入/输出)604)和/或预处理器DSP/CPU 605的输入。将这些输入模块的输出发送至一个或多个数据至尖峰编码器606。一个或多个数据至尖峰编码器606将数据编码成尖峰,所述尖峰用作模拟/混合信号神经突触阵列607的输入。

[0087] 由神经突触阵列607输出的尖峰可由尖峰译码器611解码。模式查找模块612可用于解释尖峰输出信号内的特定模式。可接着将经译码尖峰发送至一个或多个输出接口和/或中断逻辑613。配置及控制模块614允许配置及控制尖峰神经处理器的不同部分,例如设置神经突触阵列607内的突触元件及神经元的权重及配置参数。

[0088] 可根据上文所提及的实施例中的任一个对神经突触阵列607进行分段。由于电位的分级(模拟)变化,特定区段内的神经元可被动地传导,且因此仅依赖于直接且高能效的模拟计算。短暂、急剧、能量密集的动作电位经保留用于更远距离的传信,即多区段模拟/混合信号神经突触阵列内的互连遵循生物大脑的范式,即近端/近侧区段中/之间的密集连接性及远端/远侧区段之间的较不密集。因此,长程互连件610及短程互连件609两者均可存在于分段神经突触阵列607内的神经元之间。

[0089] 图6B展示互连结构的示意图。

[0090] 同样,长程互连件610及短程互连件609可存在于神经突触阵列607内的不同区段608之间。

[0091] 此(双)互连结构具有独特设计目标:对于短程通信609,近区段之间的低延迟是优先的(通过高连接密度获得),而长程通信610则以远区段之间的高吞吐量为目标。尽管区段至区段的互连结构可通过数字CMOS逻辑来实施,但神经突触阵列本身可通过使用例如模拟/混合信号、数字或非挥发性内存技术来实施。

[0092] 在其最简单的形式中,神经元薄膜电位可在模拟域中实施为电容器两端的电压,或实施为储存于数字锁存器中的多位变量。替代地,数字神经元可使用CMOS逻辑电路,例如加法器、乘法器及计数器来实现。从电路设计的角度来看,突触调节输入尖峰且将其转化为电荷,从而产生突触后电流,所述突触后电流整合于突触后神经元的薄膜电容上。硅突触的实施通常遵循混合信号方法。

[0093] 不同类型的突触传输的时空协调是推断感测信号及相关联学习性能的基础。对于较强输入,网络将自身组织成动态状态,其中神经发射主要由输入波动驱动,类似于异步-不规则状态,即时间及成对尖峰数互相关接近零。所有输入突触之间的异突触抑制在网络内产生稳定的活动序列。

[0094] 神经元产生时间锁定模式;归因于电导延迟(及可塑性规则)之间的相互作用,网络利用一组神经元群组/区段,其形成可复制且精确发射的序列。随着电导增加(因此引起对网络的净激发),发射速率可在差异系数较低的情况下增加且变得更均一。网络可在不同区段中产生具有规则时序关系(突发间间隔及占空比)的尖峰序列。

[0095] 低频同步行为的挟带包含相位重组,以使得最优(即大多数可激发)相位与在进行中的输入刺激中的事件的时间特性一致。突触电流的序列(即向外、向内)减少了在个别神经元中产生动作电位期间的时间跳动,且因此产生活动的可控性增加的网络。

[0096] 生物神经系统及常规计算在能效及认知性能方面的较大差异通过与同物理环境进行实时相互作用相关的任务进行深刻例示,尤其在存在不受控或有噪声的感测输入的情况下。然而,归因于通过实例进行学习、操作的并行性、联想记忆、多因素优化及延展性的能力,基于神经形态事件的神经网络为紧密及低功率认知系统的固有选择,所述系统学习及适应复杂感测信号的统计数据所发生的变化。

[0097] 目前用于神经形态网络的多区段可重配置架构允许为用于侦测生物医学信号的模式(例如尖峰分类、癫痫侦测等)、对图像(例如手写数字)进行分类、语音命令的应用范围提供高能效解决方案的设计,且可应用于广泛范围的装置中,包含网络物理系统及物联网中的智能型传感器或可穿戴装置。

[0098] 上述实施例中的两个或更多个可以任何适当方式组合。

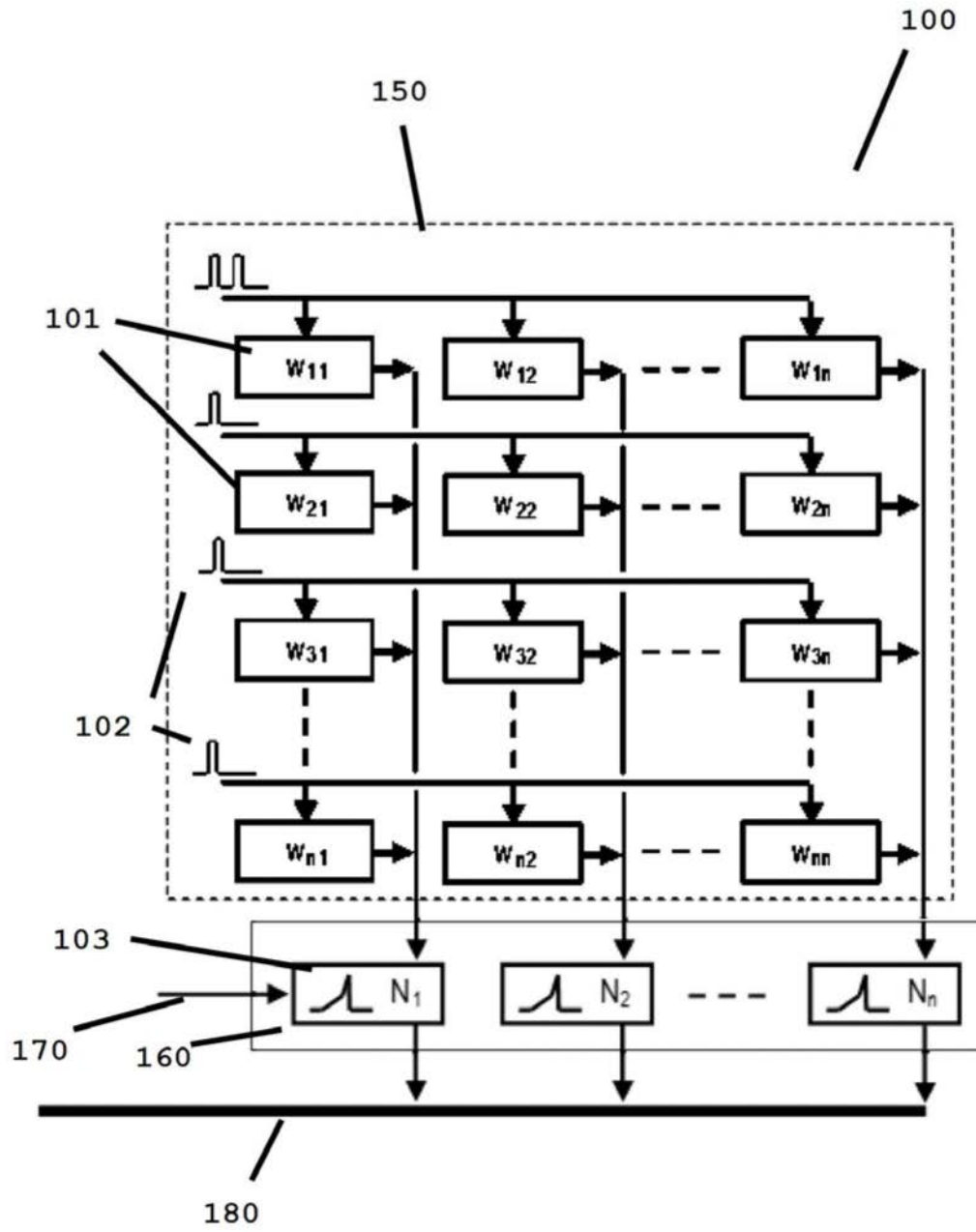


图1

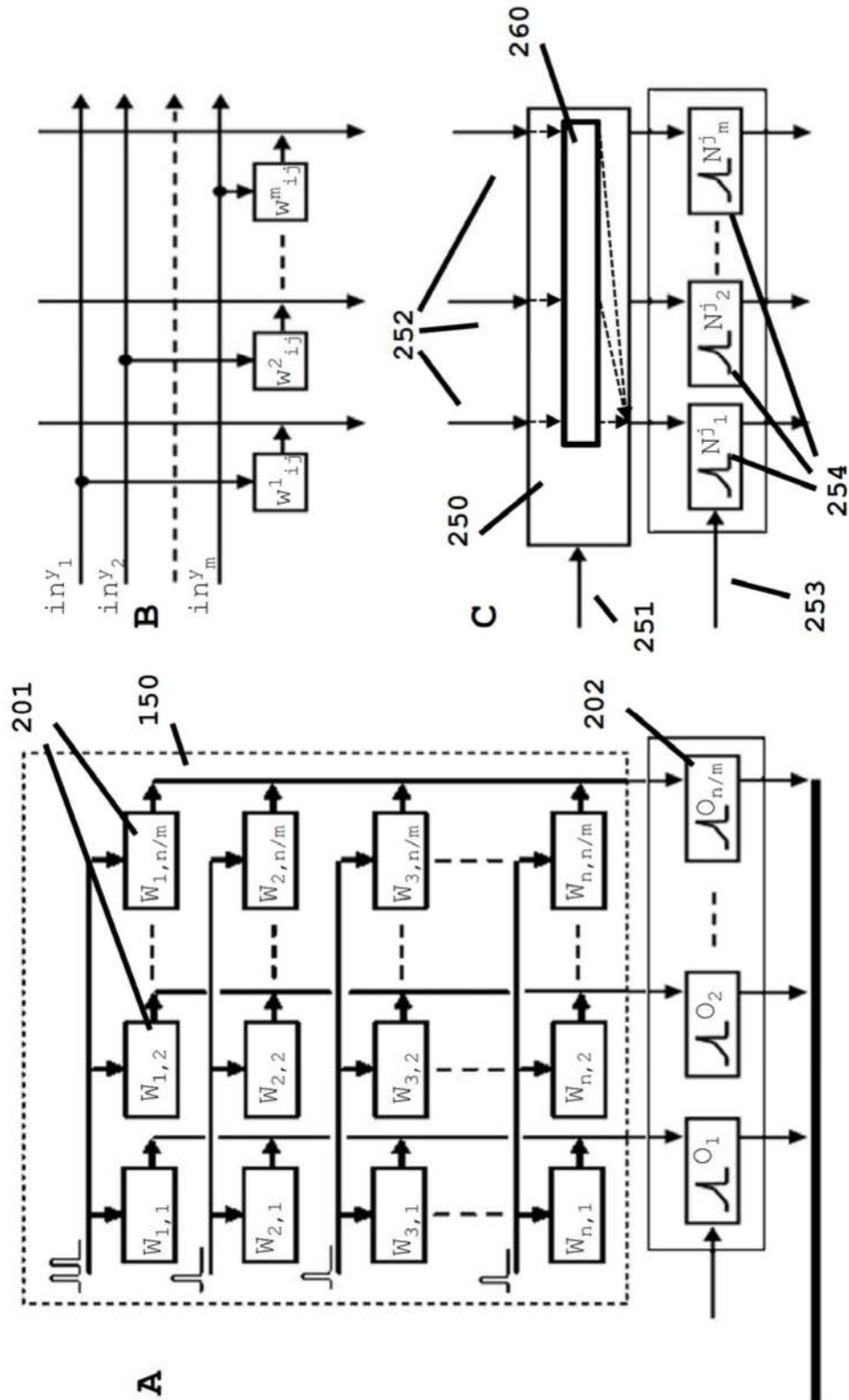


图2

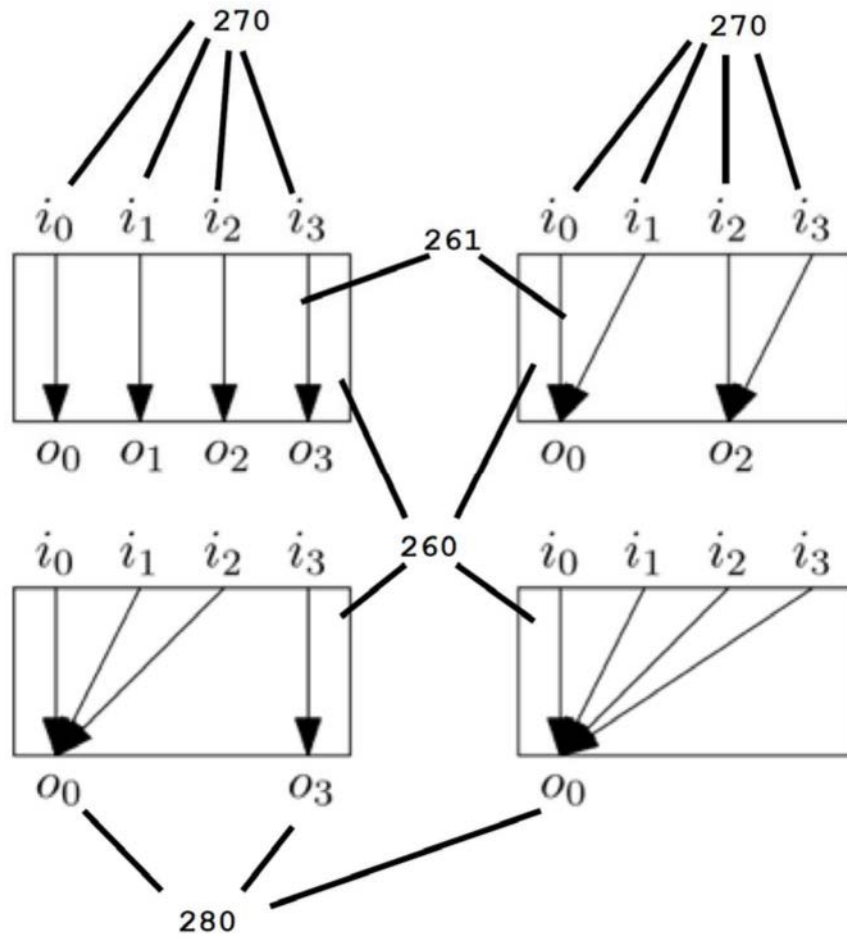


图2D

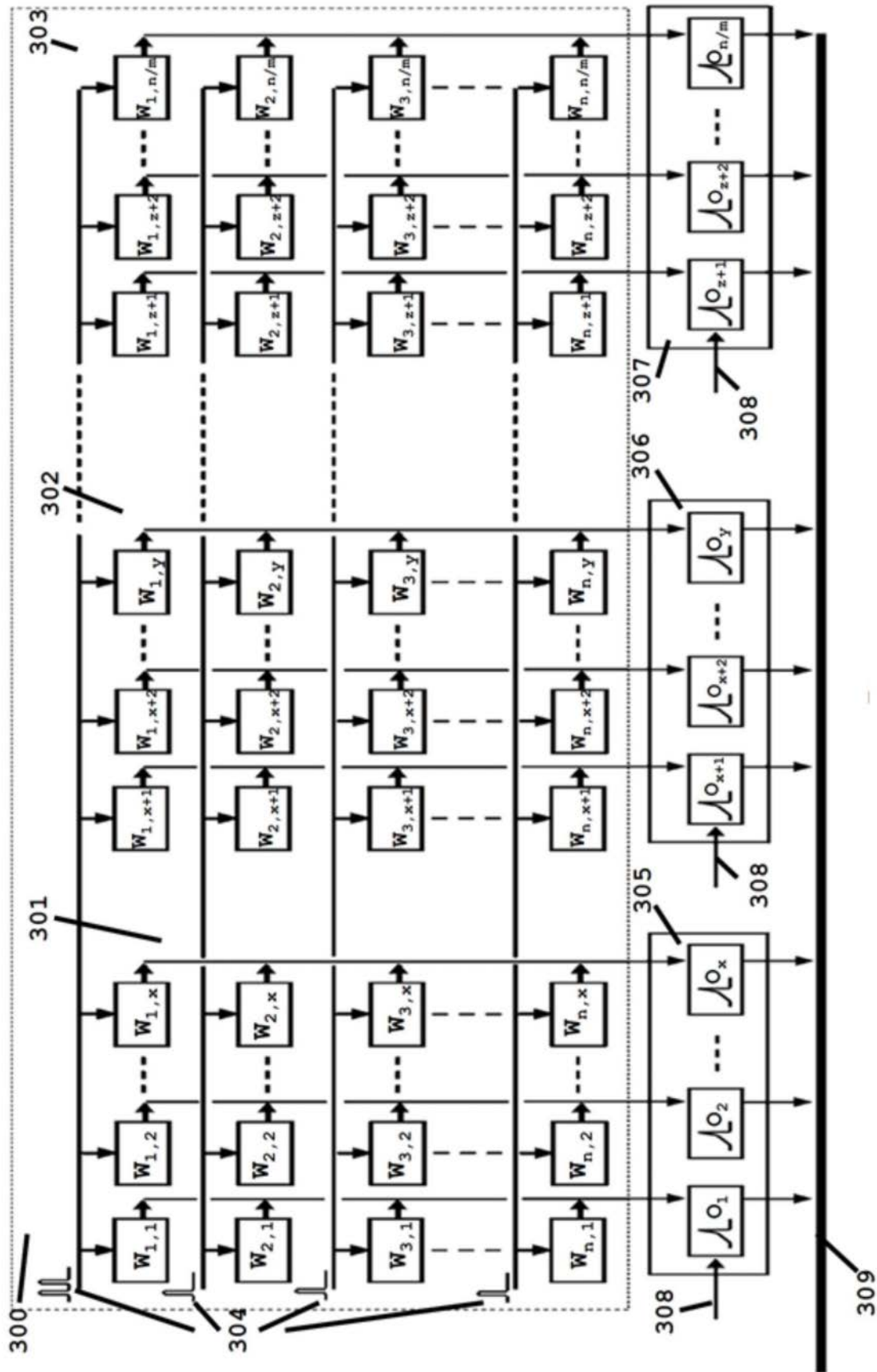


图3

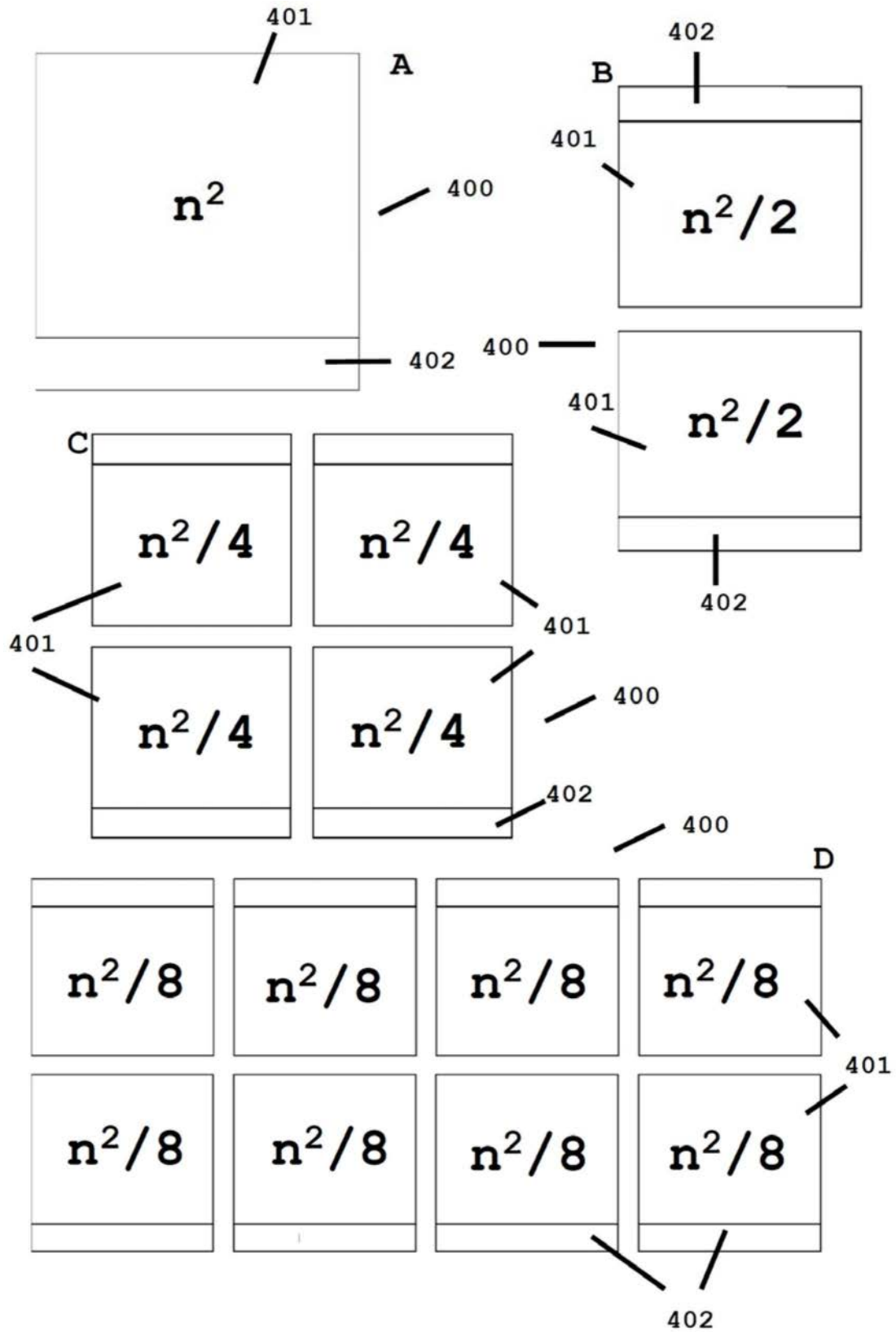


图4

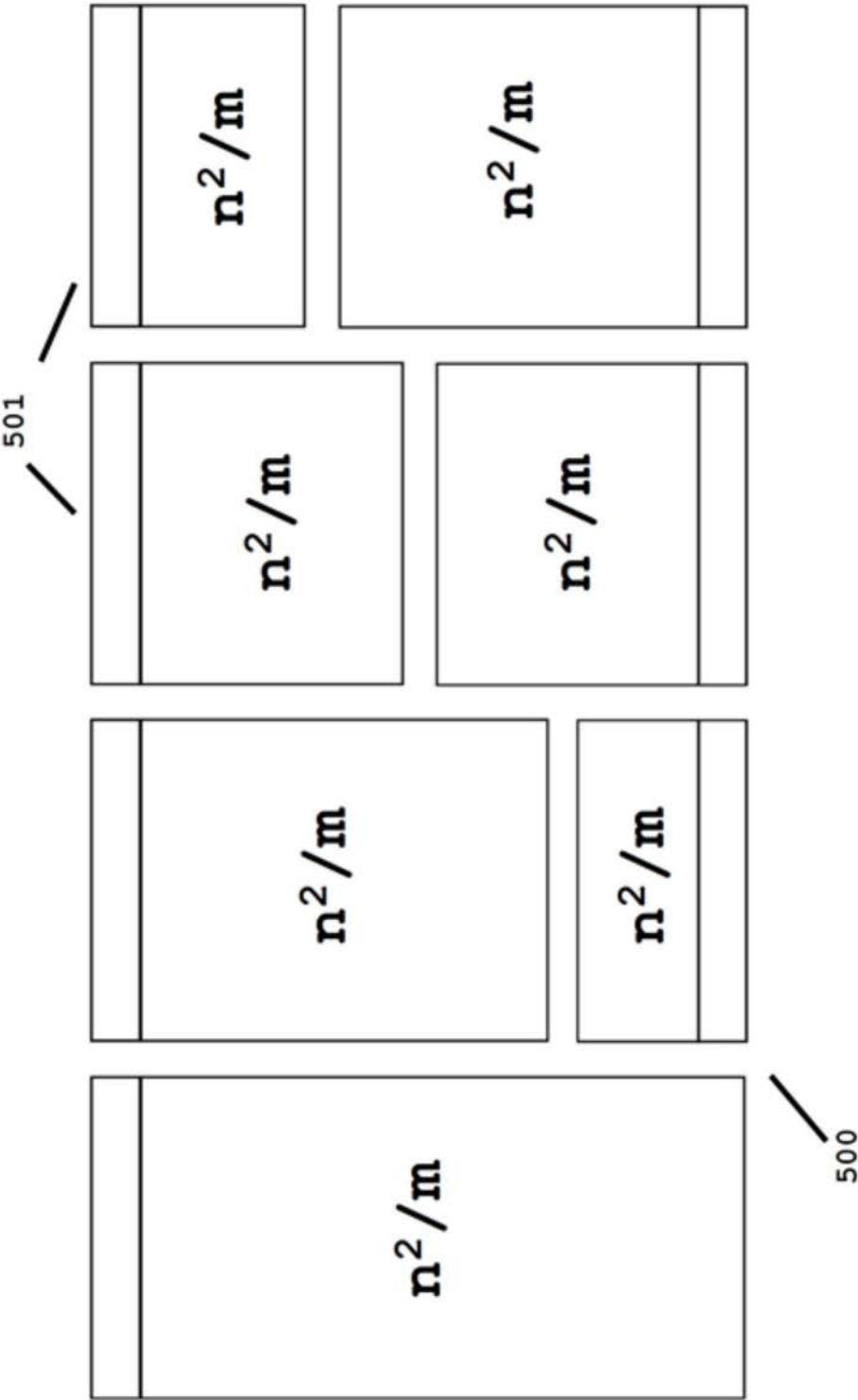


图5

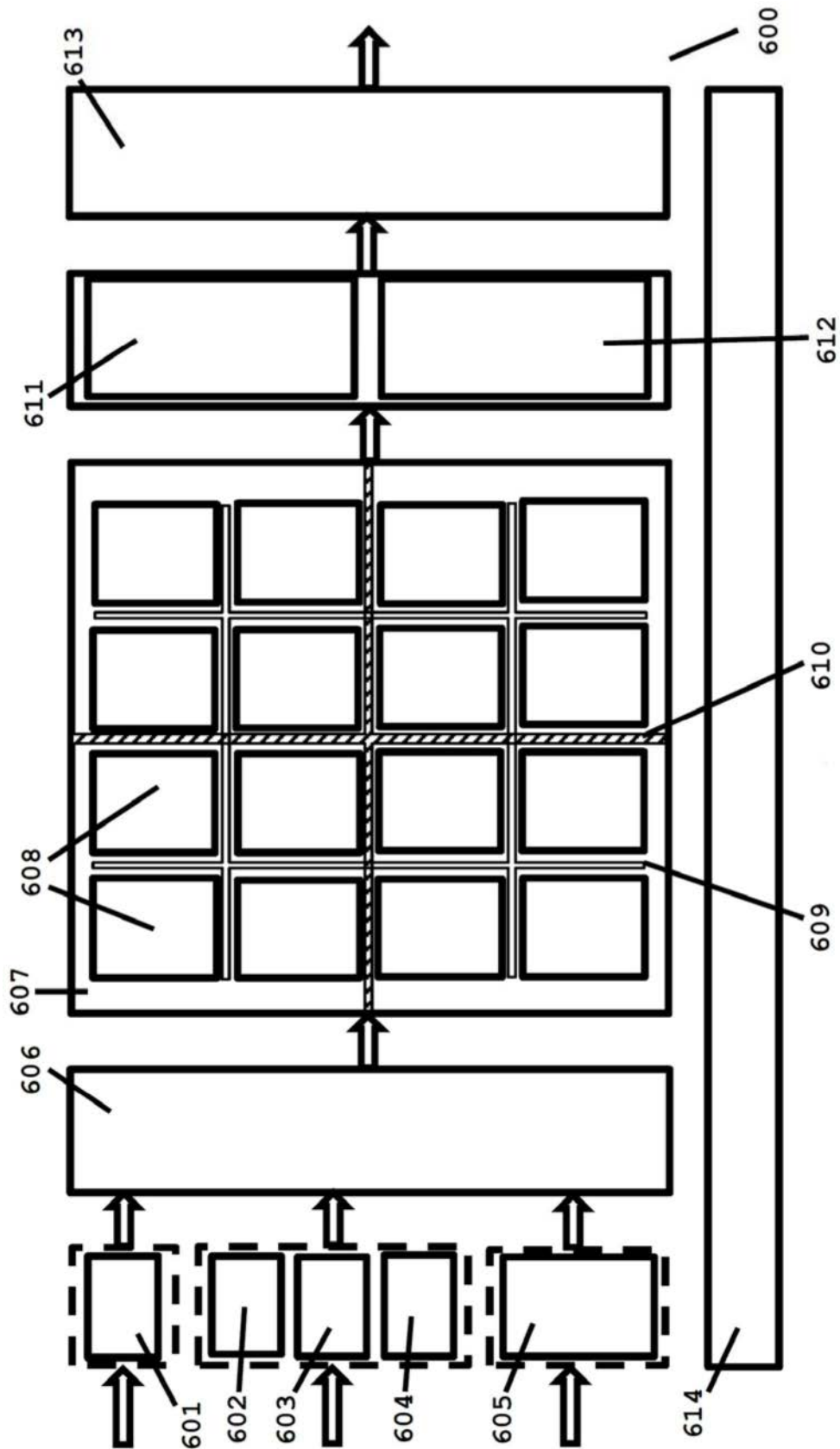


图6A

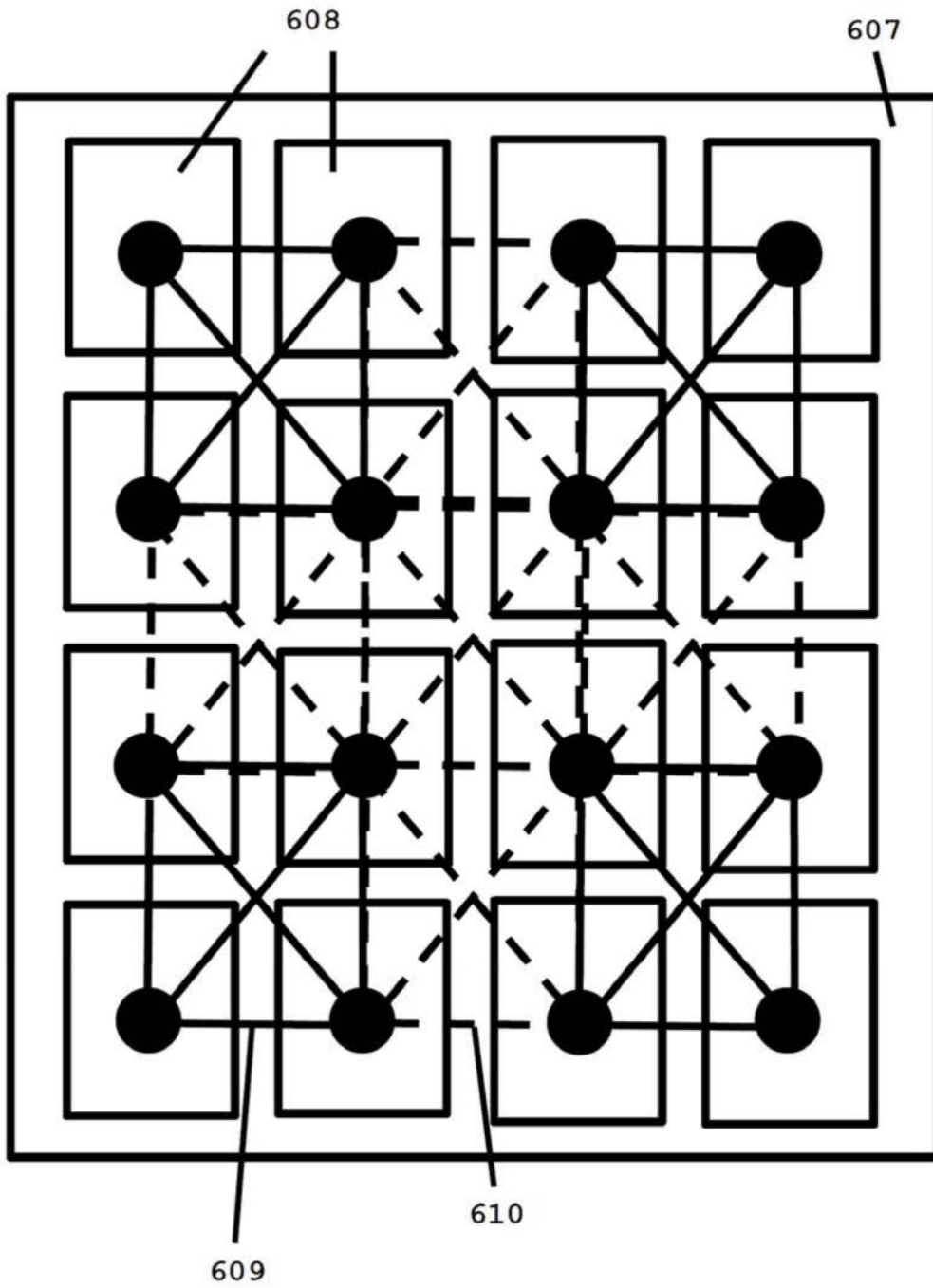


图6B